



XII Jornadas de Ingeniería Telemática

14-16 de octubre de 2015, Palma de Mallorca, Illes Balears, España



Universitat
de les Illes Balears



www.jitel.org / jitel15.uib.es

XII Jornadas de Ingeniería Telemática (JITEL 2015)

Libro de ponencias

Editores:

Maria Magdalena Payeras Capellà
Jaume Ramis Bibiloni



*Del 14 al 16 de octubre de 2015
Palma de Mallorca*



ISBN: 978-84-606-8609-5

Editores: Maria Magdalena Payeras Capellà, Jaume Ramis Bibiloni (Universitat de les Illes Balears).

El contenido de las ponencias que componen estas actas es propiedad de los autores de las mismas y está protegido por los derechos que se recogen en la Ley de Propiedad Intelectual. Los autores autorizan la edición de estas actas y su distribución a los asistentes de las XII Jornadas de Ingeniería Telemática, organizadas por la Universitat de les Illes Balears, sin que esto, en ningún caso, implique una cesión a favor de la Universitat de les Illes Balears de cualesquiera derechos de propiedad intelectual sobre los contenidos de las ponencias. Ni la Universitat de les Illes Balears, ni los editores, serán responsables de aquellos actos que vulneren los derechos de propiedad intelectual sobre estas ponencias.

© 2015, los autores

Fotografía: Francisco Jesús Valdés Campos

Portada: Francisco Jesús Valdés Campos

Maquetación: Francisca Hinarejos Campos, Jaume Ramis Bibiloni

JITEL 2015

Presentación

Nuestra sociedad está viviendo un momento de grandes avances en las tecnologías de la información y de las comunicaciones (TIC). Estos avances nos han proporcionado herramientas de comunicación para llevar a cabo las tareas más diversas. El área de ingeniería telemática se ve inmersa en esta oportunidad de progreso y es necesario que la investigación realizada en los diferentes grupos se comparta entre los investigadores del área.

Con el objetivo de servir de foro de intercambio de conocimientos para los investigadores, nacieron las Jornadas de Ingeniería Telemática (JITEL) que llegan a su duodécima edición, consolidando un foro propicio de reunión, debate y divulgación para los grupos que imparten docencia e investigan en temas relacionados con las redes y los servicios telemáticos a través del intercambio de experiencias y resultados. Conjuntamente se celebraran las IV Jornadas de Innovación Educativa (JIE) y el I Workshop QoS y QoE en Comunicación Multimedia (QQCM). Como novedad esta edición cuenta con un Fast Track para la presentación de trabajos de investigación que hayan obtenido resultados relevantes en los dos últimos años.

La edición de 2015, celebrada en Palma de Mallorca, ha sido organizada por el área de Ingeniería Telemática de la Universitat de les Illes Balears y la Asociación de Telemática (ATEL). Como imagen representativa de esta edición hemos escogido una torre vigía de la Serra de Tramuntana, levantada en el siglo XVII, que formaba parte de un sistema de comunicación mediante humos durante el día y fuegos durante la noche.

Estas actas contienen las 41 contribuciones aceptadas en Jitel, cuyas sesiones se organizan en los siguientes ámbitos temáticos: Avances en Redes de COmunicaciones móviles 5G (ARCO5G), Servicios multimedia interactivos, Seguridad, criptografía, privacidad y anonimato en las redes de comunicaciones, Gestión de datos masivos, Redes inalámbricas de área local y metropolitana, Algoritmos y Protocolos de Comunicaciones y Virtualización de red, además de dos sesiones de presentaciones breves.

Desde la organización queremos expresar nuestro agradecimiento a todos los patrocinadores y colaboradores, del evento, así como también a todos los ponentes, asistentes, miembros de los comités y revisores.

Maria Magdalena Payeras Capellà
Comité de Programa de JITEL 2015

Comité de programa

Magdalena Payeras Capellà (Universitat de les Illes Balears) - Presidenta
Eduardo Jacob Taquet (Euskal Herriko Unibertsitatea)
Javier Aracil Rico (Universidad Autónoma de Madrid)
Ignacio Soto (Universidad Carlos III de Madrid)
Luis Sánchez Fernández (Universidad Carlos III de Madrid)
Isaias Martínez Yelmo (Universidad de Alcalá)
Ramón Agüero (Universidad de Cantabria)
Amaia Méndez (Universidad de Deusto)
José Luis González Sánchez (Universidad de Extremadura)
Jesús E. Díaz Verdejo (Universidad de Granada)
Sebastián García Galán (Universidad de Jaén)
Álvaro Suárez Sarmiento (Universidad de las Palmas de Gran Canaria)
Jesús Martínez Cruz (Universidad de Málaga)
Pedro M. Ruíz (Universidad de Murcia)
Roberto García Fernández (Universidad de Oviedo)
Antonio Jose Estepa Alonso (Universidad de Sevilla)
Yannis Dimitriadis (Universidad de Valladolid)
Julián Fernández Navajas (Universidad de Zaragoza)
Javier Gozávez (Universidad Miguel Hernández de Elche)
Pilar Manzanares López (Universidad Politécnica de Cartagena)
Miguel Ángel Valero (Universidad Politécnica de Madrid)
Joaquín Salvachúa (Universidad Politécnica de Madrid)
Daniel Morató Osés (Universidad Pública de Navarra)
Jesús M. González-Barahona (Universidad Rey Juan Carlos)
Víctor M. Cameiro Díaz (Universidade da Coruña)
Jorge García Duque (Universidade de Vigo)
Santiago Felici Castell (Universitat de València)
Xavier Hesselbach Serra (Universitat Politècnica de Catalunya)
Jorge Martínez Bauset (Universitat Politècnica de València)
Miquel Oliver Riera (Universitat Pompeu Fabra)
Guiomar Corral Torruella (Universitat Ramon Llull)

Comité ejecutivo

Jesús E. Díaz Verdejo (Universidad de Granada)
Álvaro Suárez Sarmiento (Universidad de las Palmas de Gran Canaria)
Magdalena Payeras Capellà (Universitat de les Illes Balears)
Jaime Lloret Mauri (Universitat Politècnica de València)

Comité editorial

Maria Magdalena Payeras Capellà (Universitat de les Illes Balears)
Guillem Femenias Nadal (Universitat de les Illes Balears)
Ramón Agüero Calvo (Universidad de Cantabria)
Jaime Lloret Mauri (Universitat Politècnica de València)

Comité local

Andreu Pere Isem Deyà- Presidente

Felip Riera Palou

Guillem Femenias Nadal

Ignasi Furió Caldentey

Jan García Morales

Jaume Ramis Bibiloni

Javier Pastor Pérez

Josep Lluís Ferrer Gomila

Llorenç Huguet Rotger

Loren Carrasco Martorell

M. Francisca Hinarejos

Macià Mut Puigserver

Magdalena Payeras Capellà

Revisores

Isaac Agudo
Monica Aguilar
Jesus Alcober
Pablo Ameigeiras
Pablo Basanta-Val
Carlos J. Bernardos
Miguel L. Bote-Lorenzo
Fidel CACHEDA
Maria Calderon
Jesús Calle-Cancho
Maria Dolores Cano
Javier Carmona-Murillo
Vicente Casares-Giner
Baldomero Coll Perales
David Cortés-Polo
Pedro De-Las-Heras-Quiros
Isabel María Delgado Luque
Luis Diez
Gerard Draper Gil
Oscar Esparza
Diego Fernández Iglesias
Jose-Luis Ferrer-Gomila
Jesus Arias Fisteus
Marta Garcia-Arranz
Miguel Garcia-Pineda
Pablo Garrido
Jose Manuel Gimenez-Guzman
Eduardo Gómez-Sánchez
Olga León
Fidel Liberal
M. Carmen Lucas-Estañ
Jorge E. López de Vergara
Gabriel López
Gabriel Macia-Fernandez
Elsa Macias Lopez
Rafa Marin-Lopez
Domingo Marrero Marrero
Sergio Montero
Jose Ignacio Moreno Novella
Macià Mut Puigserver
Juan Pedro Muñoz-Gea
Jorge Navarro-Ortiz
Ana Nieto
Francisco J. Novoa
Toni Oller Arcas
Laura Pozueco
Francisco Javier Rodriguez-Perez
Jose Luis Salazar
Juan Carlos Sanchez-Aarnoutse
Jose Santana
Roberto Sanz
Jaume Segura Garcia
Miguel Sepulcre
Federico Simmross-Wattenberg
Antonio Skarmeta Gomez
Manuel Urueña Pascual
Eduardo Jacob Taquet
Javier Aracil Rico
Ignacio Soto
Luis Sánchez Fernández
Isaias Martinez Yelmo
Ramón Agüero
Amaia Méndez
José Luis González Sánchez
Jesús E. Díaz Verdejo
Sebastián García Galán
Álvaro Suárez Sarmiento
Jesús Martínez Cruz
Pedro M. Ruíz
Roberto García Fernández
Antonio Jose Estepa Alonso
Yannis Dimitriadis
Julián Fernández Navajas
Javier Gozávez
Pilar Manzanares López
Miguel Ángel Valero
Joaquín Salvachúa
Daniel Morató Osés
Jesús M. González-Barahona
Víctor M. Cameiro Díaz
Jorge García Duque
Santiago Felici Castell
Xavier Hesselbach Serra
Jorge Martínez Bauset
Miquel Oliver Riera
Guiomar Corral Torruella

QQCM 2015

Presentación

Internet se encuentra en una encrucijada, sus usuarios se han acostumbrado a la rápida implantación de acceso a contenidos multimedia, transferencia de información multimedia entre usuarios particulares, videoconferencias, juegos online... La ingente cantidad de información personalizada y los servicios disponibles crecen exponencialmente, pero los consumidores se ven atados a una red best effort que ofrece una respuesta muy inestable. Proporcionar nuevos servicios multimedia a través de Internet permite una gran flexibilidad, pero la calidad ofrecida puede ser bastante pobre. Por eso es necesario un análisis de la calidad ofrecida por la red, QoS (Quality of Service) y la calidad percibida por el usuario del servicio, QoE (Quality of Experience).

En esta línea, el presente workshop abarca los siguientes tópicos de interés: repaso de las principales arquitecturas desarrolladas para asegurar la QoS en Internet y si están preparadas para soportar la implantación masiva de nuevos servicios multimedia. Propuesta de nuevos protocolos de comunicación para los servicios multimedia y evaluación de los mismos. Análisis de cómo la calidad percibida está en consonancia con los parámetros de QoS asegurados por las redes para servicios multimedia. Propuesta de nuevos parámetros adicionales (facilidad de manejo, receptividad, preferencias...) que introducen el factor humano y definen precisamente el consumo del servicio multimedia o la forma en que el usuario interactúa con el mismo.

Julián Fernández Navajas

Comité de Programa de QQCM 2015

Comité de programa

Julián Fernández Navajas (Universidad de Zaragoza) - Presidente
Armando Ferro Vázquez (Universidad del País Vasco)
María Ángeles Vázquez Castro (Universidad autónoma de Barcelona)
Jorge Navarro Ortiz (Universidad de Granada)
Álvaro Suárez Sarmiento (Universidad de Las Palmas de Gran Canaria)
Matías Toril Genovés (Universidad de Málaga)
Mónica Aguilar Igartua (Universidad Politécnica de Cataluña)
María Dolores Cano Baños (Universidad Politécnica de Cartagena)
Juan Carlos Guerri Cebollada (Universidad Politécnica de Valencia)
Jaime Lloret Mauri (Universidad Politécnica de Valencia)

JIE 2015

Presentación

Se recogen en estas actas los trabajos presentados en las IV Jornadas de Innovación Educativa en Ingeniería Telemática (JIE), celebradas en Palma de Mallorca, del 14 al 16 de octubre de 2015. Se consolidan así estas Jornadas, que se iniciaron coincidiendo con la adaptación al Espacio Europeo de Educación Superior (EEES), lo que supuso, y sigue suponiendo, un verdadero desafío para la comunidad docente universitaria.

Al igual que en las ediciones anteriores, celebradas en Valladolid, Santander y Granada, estas Jornadas se presentan como una oportunidad para crear un foro de encuentro e intercambio de experiencias de innovación docente entre profesores en el ámbito de la Ingeniería Telemática. Con este objetivo las sesiones de la presente edición se han estructurado de manera que abarcan como principales tópicos de interés el seguimiento, tutorización y evaluación de los estudiantes, el uso y aplicación de las tecnologías de la información y las comunicaciones en la docencia universitaria en general y en la docencia de la telemática en particular y, para finalizar, las metodologías docentes y organización y conducción de las clases.

Los trabajos presentados para su posible aceptación en estas Jornadas se sometieron a un riguroso proceso de revisión por pares, en el que cada trabajo fue evaluado por dos revisores, garantizando así, bajo la supervisión del Comité de Programa, la originalidad y la calidad de su contenido. En estas actas se recogen las nueve contribuciones que fueron finalmente aceptadas para su presentación en dichas Jornadas. Estas presentaciones se han estructurado en 2 sesiones desarrolladas conjuntamente con las XII Jornadas de Ingeniería Telemática (JITEL 2015).

Para concluir, quisiera agradecer al Comité de Programa su colaboración en todas las tareas organizativas de estas Jornadas, así como a los revisores su inestimable trabajo. Sin su participación no hubiera sido posible llevar a cabo estas Jornadas que, edición tras edición, se consolidan como un referente en la innovación educativa del área de Ingeniería Telemática.

Sed bienvenidos y disfrutad de las Jornadas.

Jaume Ramis Bibiloni

Comité de Programa de JIE 2015

Comité de programa

Jaume Ramis Bibiloni (Universitat de les Illes Balears)- Presidente
María Jesús Verdú Pérez (Universidad de Valladolid)
Raquel Crespo García (Universidad Carlos III de Madrid)
José Manuel Giménez Guzmán (Universidad de Alcalá)
Iván Marsá Maestre (Universidad de Alcalá)
Ramón Agüero Calvo (Universidad de Cantabria)
Alberto Eloy García Gutiérrez (Universidad de Cantabria)
José Ángel Irastorza Teja (Universidad de Cantabria)
Jorge Navarro Ortiz (Universidad de Granada)
Elsa Macías López (Universidad de las Palmas de Gran Canaria)
Jesús Martínez Cruz (Universidad de Málaga)
Yannis Dimitriadis (Universidad de Valladolid)
Martín Llamas Nistal (Universidad de Vigo)
Guillermo Azuara Guillén (Universidad de Zaragoza)
Fidel Cacheda Seijo (Universidade da Coruña)
Diego Fernández (Universidade da Coruña)
Francisco Barceló Arroyo (Universitat Politècnica de Catalunya)

Revisores

María Jesús Verdú Pérez
Raquel Crespo García
José Manuel Giménez Guzmán
Iván Marsá Maestre
Ramón Agüero Calvo
Alberto Eloy García Gutiérrez
José Ángel Irastorza Teja
Jorge Navarro Ortiz
Elsa Macías López
Jesús Martínez Cruz
Yannis Dimitriadis
Martín Llamas Nistal
Guillermo Azuara Guillén
Fidel Cacheda Seijo
Diego Fernández
Francisco Barceló Arroyo
Luisa Regueras

Lista de Contribuciones

Fast Track

Estimación del RSSI en WiFi Teniendo en Cuenta Restricciones Prácticas	1
<i>Elsa Macías López, José A. Santana, Domingo Marrero, Vicente Mena, Álvaro Suárez</i>	
All-Path Bridging: Path-exploration VS Path-computation on data centers and campus networks	2
<i>Elisa Rojas, Guillermo Ibáñez, José Manuel Giménez-Guzmán, Juan A. Carral, Alberto García-Martínez, Isaias Martínez-Yelmo, José Manuel Arco</i>	
Diseño de una arquitectura para la red de acceso para 5G basada en SDN	3
<i>Pablo Ameigeiras, Juan J. Ramos-Muñoz, Laurent Schumacher, Jonathan Prados-Garzón, Jorge Navarro-Ortiz, Juan M. López-Soler</i>	
Design of a Synchronous Ethernet in a SDN architecture	4
<i>Raúl Suárez, David Rincón, Sebastià Sallent</i>	
Emerging Real-time Services: Optimizing Traffic by Smart Cooperation in the Network	5
<i>Julián Fernández Navajas, José Ruiz Mas, José María Saldaña Medina, Dan Wing, Muthu A. M. Perumal, Michael Ramalho, Gonzalo Camarillo, Fernando Pascual, Diego R. López, Manuel Núñez, David Flórez, Juan A. Castell, Tomaso de Cola, Matteo Berioli</i>	
Recuperación de Información desde diferentes perspectivas	6
<i>Diego Fernández, Víctor Carneiro, Francisco Nóvoa de Manuel, Xacobe Macía da Silva, Fidel Cacheda</i>	
Monitorización Estadística Multivariante para Ciberseguridad: Sistema VERITAS	7
<i>José Camacho, Gabriel Maciá Fernández, Pedro García Teodoro</i>	

Jitel 1: Avances en Redes de Comunicaciones móviles 5G (ARCO5G)

Joint Strategy for LTE Resource Allocation: Multicast Subgrouping & Unicast Transmissions	9
<i>Alejandro de La Fuente, Raquel Pérez Leal and Ana García Armada</i>	
Asignación de Recursos Radio en Redes Industriales Inalámbricas Multisalto	16
<i>Sergio Montero, Javier Gozávez and Miguel Sepulcre</i>	
Effect of Network Coding and Multi-hop Beaconing on the Channel Load of Vehicular Networks	24
<i>Miguel Sepulcre, Javier Gozávez, M. Carmen Lucas-Estañ, Baldomero Coll-Perales, Sergio Montero, José Ramón Gisbert and Alejandro Moraleda</i>	
Downlink Optimisation of FFR-Aided OFDMA Networks using Channel-Aware Scheduling	30
<i>Jan García-Morales, Guillem Femenias and Felip Riera-Palou</i>	
Propuesta de un algoritmo de encaminamiento para distribuir la exposición a los campos electromagnéticos en redes inalámbricas malladas	38
<i>Julián Igareda, Luis Díez and Ramón Agüero</i>	
Análisis del Rendimiento de Técnicas de Network Coding y Encaminamiento Oportunista Sobre Redes Malladas Inalámbricas	44
<i>Pablo Garrido, David Gómez Fernández, Ramón Agüero and Luis Muñoz</i>	

Performance evaluation of dissemination protocols for emergency messages in Vehicular Ad-Hoc networks	52
<i>Cristhian Iza Paredes, Ahmad Mohamad Mezher and Mónica Aguilar Igartua</i>	

Jitel 2: Short Papers I

Certificado Digital Accesible para la e-Administración	60
<i>Boni García, Ana Gómez, Rafael Conde, Yolanda Hernández and Miguel Ángel Valero</i>	
AC TV Manager: Nuevo sistema de gestión para IPTV basado en software libre	68
<i>Alain Chas, Jose M. Jiménez, Andrés López-Herrerros, Alejandro Canovas, Jaime Lloret</i>	
Diseño SDN/NFV con retardo balanceado entre orquestador, controladores y elementos de red	76
<i>Aurelio J. García, Cristina Cervelló-Pastor and Yury A. Jiménez</i>	
Transmisión de video H.264/AVC con Adaptive Bitrate Streaming y control de tasa en el servidor	82
<i>Jorge Mata, Edwin Solorzano, Jose L. Muñoz, Juan Hernández, Juanjo Alins and Oscar Esparza</i>	
Estudio sobre necesidades, hábitos, preferencias y expectativas de los usuarios con respecto a la sincronización multimedia en escenarios híbridos	88
<i>Fernando Boronat, Mario Montagud, Marc Martínez and Dani Marfil</i>	

Jitel 3: Short Papers II

Herramientas TIC para la mejora del rendimiento y la planificación de redes PLC para Smart Grids	96
<i>Miguel Seijo Simó, Gregorio López López, José Ignacio Moreno Novella, Javier Matanza Domingo, Fernando Martín Soto and Cristina Martínez Ruiz</i>	
V2I en campo cercano: modelado de las comunicaciones Eurobaliza Tren para análisis de interfe- rencias a sistemas de señalización ferroviaria	104
<i>Lara Rodríguez, Christian Pinedo, Igor López, Marina Aguado, Jasone Astorga, Mariví Higuero, Iñigo Adin and Guillermo Bistue</i>	
Adherencia al tratamiento de deformidades craneales con un dispositivo ortopédico sensorizado	110
<i>Lucero Rosas, Julio García, Antoni Oller, Juan López, and Jesús Alcober</i>	
Influence of the path establishment method on an SDN controller time response	116
<i>David Quiroz and Cristina Cervelló-Pastor</i>	
Acceso automático a información museística en dispositivos móviles	124
<i>Guillermo Azuara, Ana María López, Ana Salinas, Eduardo Pascual and Jaime Vicente</i>	
Modelo para la ubicación de servicios en Cloud Híbrido	130
<i>Adrià Mallorquí, Agustín Zaballos, Joan Navarro, Ramón Martín de Pozuelo and Alan Briones</i>	

Jitel 4: Servicios multimedia interactivos

An objective video quality metric for live video streaming over cloud systems: full and non reference approaches	138
<i>Jaume Segura-García, Santiago Felici-Castell, Miguel García-Pineda, Enrique Chirivella-Pérez, Juan Gutiérrez-Aguado and José M Claver Iborra</i>	
Evaluation of the Trade-off between Energy Consumption and QoE in Video Transmission with MANET Routing Protocols	145
<i>Ramón Sánchez-Iborra and María-Dolores Cano</i>	
Testing Infrastructure for WebRTC applications	153
<i>Boni García, Luis López-Fernández, Micael Gallego, Francisco Gortázar and Francisco Javier López</i>	

Maqueta para la evaluación de la calidad de experiencia del servicio videostreaming 3D en LTE . . .	161
<i>Inmaculada Palomo, Marta Solera-Delgado, Matías Toril, Javier Poncela-González and Gerardo Gómez</i>	
Jitel 5: Seguridad, criptografía, codificación, privacidad y anonimato en las redes de comunicaciones	
Problemas de seguridad y amenazas que pueden afectar al cloud híbrido de FINESCE	169
<i>Julia Sánchez, Guiomar Corral, Jaume Abella and Agustín Zaballos</i>	
Two Approaches to Obtain Anonymity in Transferable Electronic Ticketing Schemes	177
<i>Macià Mut Puigserver, Jordi Castellà Roca and M. Magdalena Payeras Capellà</i>	
Comparación de modelos de procedimiento en el análisis forense digital	185
<i>Jordi Blanco-Galdeano, Helena Taribó-Gómez, Juan Vera-Del-Campo and Miguel Soriano</i>	
Jitel 6: Gestión de datos masivos	
Primeros resultados hacia la detección automática de bots en Twitter	191
<i>Luis Sánchez Fernández, Tello Miñana Rontomé, Jesús Arias Fisteus, Pablo Basanta-Val, Damaris Fuentes-Lorenzo, M. Luz Congosto and Norberto Fernández García</i>	
Análisis de Datos Funcionales para Gestión de Red: Técnicas, Retos y Oportunidades	197
<i>David Muelas, Jorge E. López de Vergara, Jose R. Berrendero, Javier Aracil</i>	
Ejecución de JavaScript en Paralelo en Secuencias de Navegación Web	205
<i>Jose Losada, Juan Raposo, Alberto Pan, Paula Montoto and Manuel Álvarez</i>	
Monitorización y Análisis de Tráfico de Red con Apache Hadoop	213
<i>Rubén García-Valcárcel, Rafael Leira, Iván González and Jorge E. López de Vergara</i>	
Jitel 7: Redes inalámbricas de área local y metropolitana	
Uso de información de la posición de los nodos en el protocolo de enrutamiento DYMO	221
<i>Enrica Zola, Francisco Barceló-Arroyo and Israel Martín-Escalona</i>	
Uso de OpenFlow para la gestión de caminos redundantes en redes inalámbricas	226
<i>Elías Molina, Eduardo Jacob and Armando Astarloa</i>	
Uso de técnicas de emulación sobre la plataforma ns-3 para el análisis de QoS de servicios multimedia	232
<i>Juan Arroyo, Luis Díez and Ramón Agüero</i>	
Uso de técnicas de negociación automática para la asignación de canales en IEEE 802.11	239
<i>Enrique De La Hoz De La Hoz, Jose Manuel Gimenez-Guzman, Iván Marsá-Maestre and David Orden</i>	
Estimación Eficiente del RSSI en Redes WiFi para Servicios Inalámbricos Futuros	247
<i>José A. Santana, Elsa Macías López, Vicente Mena, Domingo Marrero and Álvaro Suárez</i>	
Análisis de las prestaciones de un esquema de codificación lineal aleatoria sobre redes inalámbricas	255
<i>Pablo Garrido Ortiz, David Gómez Fernández, Jorge Lanza Calderón, Ramón Agüero and Joan Serrat</i>	
Jitel 8: Algoritmos y Protocolos de Comunicaciones. Virtualización de red	
A Dynamic Multimetric Weights Distribution in a Multipath Routing Protocol using Video-Streaming Services over MANETs	263

<i>Ahmad Mohamad Mezher, Cristhian Iza Paredes, Carolina Tripp-Barba and Mónica Aguilar Igartua</i>	
Joint impact of CQI feedback delay and CQI reporting rate on channel-aware scheduling	271
<i>José Óscar Fajardo, Ianire Taboada and Fidel Liberal</i>	
REMODE: Un algoritmo para el cálculo de la reputación de nodos y caminos en redes ad hoc cooperativas	277
<i>Lenin Bladimir Guaya Delgado, Esteve Pallarés Segarra and Jordi Forné Muñoz</i>	
Multi constraint network virtualization strategy based on Paths Algebra for non-linear parameters	285
<i>Xavier Hesselbach, José Roberto Amazonas and Alejandro Gutiérrez</i>	
Experimental evaluation of mobility management protocols for multimedia communications	293
<i>Javier Carmona-Murillo, David Cortés-Polo, Jesús Calle-Cancho, José-Luis González-Sánchez and Francisco Javier Rodríguez-Pérez</i>	

QoS y QoE en Comunicación Multimedia (QQCM)

Optimización y securización de servicios multimedia interactivos: impacto en las relaciones comerciales online	301
<i>Julián Fernández-Navajas and José Saldana</i>	
Nuevas arquitecturas para la adaptación de servicios multimedia en redes móviles emergentes	309
<i>José Óscar Fajardo, Ianire Taboada, José Ramón Otegi, Eva Ibarrola, Armando Ferro and Fidel Liberal</i>	
Wireless and Multimedia Networking (WiMuNet) Laboratory	315
<i>Jorge Navarro-Ortiz, Pablo Ameigeiras, Juan J. Ramos-Munoz, Jonathan Prados-Garzón and Juan M. López-Soler</i>	
Maqueta para la evaluación de la calidad de experiencia del servicio videostreaming 3D en LTE	323
<i>Inmaculada Palomo, Marta Solera Delgado, Matías Toril, Javier Poncela-González, Gerardo Gómez</i>	
Design of smart services and routing protocols for VANETs in smart cities	331
<i>Ahmad Mohamad Mezher, Cristhian Iza Paredes, Luis Urquiza Aguiar, Ángel Torres Moreira and Mónica Aguilar Igartua</i>	
QoE en redes adhoc, descarga adaptativa de contenidos y vídeo 3D	339
<i>Pau Arce, Ismael de Fez, Francisco Fraile, Santiago González, Paola Guzmán and Juan C. Guerri</i>	
Estimación Proactiva de QoS de Canales Inalámbricos para Mejorar la QoE en Servicios Multimedia	347
<i>Álvaro Suárez, Elsa Macías, Vicente Mena, Domingo Marrero and José Aurelio Santana</i>	
Impacto de Parámetros de QoS en Aspectos de QoE: Análisis desde el Punto de Vista de la Sincronización Multimedia	355
<i>Mario Montagud, Fernando Boronat, Marc Martínez, Jordi Belda and Pablo Cesar</i>	
Estudio de QoS y QoE y Propuestas de Sistemas para la Mejora de los Servicios Multimedia e IPTV	363
<i>Jose M. Jiménez, Alejandro Cánovas, Juan R. Díaz and Jaime Lloret</i>	

Jie 1: Evaluación y seguimiento en la docencia / TICs aplicadas a la docencia

Evaluación para el aprendizaje aplicando revisión entre iguales: caso de estudio en una asignatura de programación	372
<i>Raquel M. Crespo García, Iria Estévez-Ayres and Carlos Delgado Kloos</i>	
Coevaluación de trabajos mediante el empleo de una aplicación móvil	378
<i>José Luis Tornos, José Luis Salazar, Guillermo Azuara, M^a Gloria Mendieta and Pilar Del Valle</i>	
Implementación de una solución VDI en laboratorios docentes universitarios	383

Xacobe Macía Da Silva, Víctor Manuel Carneiro Díaz, Fidel Cacheda Seijó, Francisco Javier Nóvoa de Manuel and Diego Fernández Iglesias

Diseñando un MOOC en edX: Introducción a la Programación con Java – Parte 1 391

Carlos Alario-Hoyos, Carlos Delgado Kloos, Iria Estévez-Ayres, Carmen Fernández-Panadero, Jorge Blasco, Sergio Pastrana, Guillermo Suárez-Tangil and Julio Villena-Román

Simtools: Teaching Networks with Virtual Scenarios 399

Juanjo Alins, José L. Muñoz, Jorge Mata and Oscar Esparza

Jie 2: Metodologías docentes y organización

Propuesta Metodológica para Asignaturas de Telemática en Máster Profesionales Oficiales 405

Elsa Macías Lopez, Álvaro Suárez Sarmiento, Jose Aurelio Santana, Domingo Marrero Marrero and Vicente Mena

Una experiencia de "Flipped Classroom" en Ingeniería Telemática 413

Iván Marsá Maestre, Juan R. Velasco, Enrique De La Hoz and José Manuel Giménez-Guzmán

LIDER: experiencia para mejorar la eficiencia del aprendizaje colaborativo en grupos de tamaño medio 420

Guillermo Azuara, María Carmen Aguilar, Julián Fernández-Navajas, José Luis Salazar, Inmaculada Plaza and José Luis Tornos

Gamificación en el aula universitaria: un caso práctico en una asignatura de Redes 426

Diego Fernández Iglesias, Fidel Cacheda Seijó, Francisco Javier Nóvoa de Manuel, Xacobe Macía Da Silva and Víctor Carneiro Díaz

Fast Track

Estimación del RSSI en WiFi Teniendo en Cuenta Restricciones Prácticas

Elsa Macías López^{2,3}, José A. Santana^{1,3}, Domingo Marrero^{2,3}, Vicente Mena^{1,3}, Álvaro Suárez^{2,3}

¹Departamento de Señales y Comunicaciones, {jose.santana,vicenteefigenio.mena}@ulpgc.es

²Departamento de Ingeniería Telemática, {alvaro.suarez, elsa.macias, domingo.marrero}@ulpgc.es

³Grupo de Arquitectura y Concurrencia (GAC)

Universidad de Las Palmas de Gran Canaria

Campus Universitario de Tafira s/n

35017 Las Palmas de Gran Canaria

Resumen- Este esquema está basado en [1] y además presentamos algunas reflexiones nuevas de cara a presentar los avances que hemos logrado en el Grupo en los últimos años.

Palabras Clave- RSSI, Wireless Fidelity, Estimación, Medias Móviles, Aplicaciones Inalámbricas.

I. INTRODUCCIÓN

Received Signal Strength Indicator (RSSI) indica el nivel de potencia de señal que recibe un dispositivo inalámbrico. Una gran cantidad de servicios inalámbricos utilizan una estimación de esta medida. Ejemplos de estos servicios son: estimación de la distancia a la que se encuentra el dispositivo de un *Punto de Acceso (PA)*, mecanismos de localización mediante triangulación de al menos tres PA, transmisión adaptativa y control de consumo energético, estimación de la calidad del enlace inalámbrico, indicación del nivel de cobertura radio respecto a un conjunto de PA, controlar la primera fase de fisgoneo (*scanning*) del *roaming*... Nosotros lo hemos utilizado para controlar interrupciones de servicio en sesiones de video Streaming inalámbrico [2].

Estimar eficientemente el RSSI es muy importante: De ello depende que los servicios inalámbricos funcionen eficientemente.

II. LAS RESTRICCIONES PRÁCTICAS

En el último año, se han publicado métodos de estimación del RSSI en revistas muy importantes. En [3] se demuestra un hecho revelador importante: la variación del RSSI a lo largo del tiempo tiende a definir medias móviles. Algo que nosotros habíamos aprovechado sin demostrar formalmente desde mucho antes [4]. Trabajo que se ha mejorado recientemente en [5]. Ninguno de los trabajos anteriores había tenido en cuenta restricciones reales de la estimación del RSSI en ambientes reales.

Para empezar, el RSSI es producido por una etapa Hardware dentro de las tarjetas de comunicación *Wireless Fidelity (WiFi)*. Estas tarjetas producen valores cuyos máximos y mínimos no se dan a conocer por parte del fabricante. Por tanto, cada tarjeta de red exporta sus propios valores. Estos valores los trata el *driver* que permite que sean leídos por el desarrollador de aplicaciones a través de las llamadas al sistema de los sistemas operativos de los dispositivos. Los mecanismos que permite el sistema

operativo para leer estos valores son muy variados, por ejemplo en los teléfonos móviles suelen haber restricciones para acceder a estos valores mientras que en los equipos de sobremesa suele haber más facilidad, y en los *ebook* hay que buscar software especializado para leerlos. En definitiva, el software que permite leer estos valores suele incluir sus propias restricciones.

En la tabla 1 de [1] mostramos 8 combinaciones usando: Windows, Linux, Android, con *Netstumbler* y *tcpdump* para medir y probar diferentes estimadores de RSSI. Con estas pruebas demostramos que los resultados óptimos de [3] y [5] no eran tan buenos como con los resultados de laboratorio o sintéticos con los que fueron probados.

III. EL NUEVO ESTIMADOR

Para solventar los problemas anteriores diseñamos un nuevo estimador (apartado 4 de [1]) basado en [1] pero que era capaz de comportarse mejor en varias situaciones reales como se mostró en el apartado 5 de [1].

El uso de este estimador a nivel de aplicación es importante para los nuevos servicios telemáticos inalámbricos dada la variedad de dispositivos y equipos inalámbricos que existen en el Mercado. Y es especialmente interesante en el marco de las *Software Defined Networks* porque interesa definir funciones virtualizadas que entreguen valores de RSSI que puedan ser estimados de forma general.

REFERENCIAS

- [1] A. Suárez, J.A. Santana, E.M. Macias-Lopez, V.E. Mena, J.M. Canino, and D. Marrero, "RSSI Prediction in WiFi Considering Realistic Heterogeneous Restrictions," *Network Protocols and Algorithms*, vol. 6, 2014, pp. 19–40.
- [2] Á. Suárez, M. La-Menza, E.M. Macías, and V.S. Sunderam, "Automatic Resumption of Streaming Sessions over Wireless Communications Using Agents.," *IMECS*, 2006, pp. 926–931.
- [3] E. Chin, D. Chieng, V. Teh, M. Natkaniec, K. Loziak, and J. Gozdecki, "Wireless link prediction and triggering using modified Ornstein–Uhlenbeck jump diffusion process," *Wireless Networks*, vol. 20, 2014, pp. 379–396.
- [4] A. Suárez, K.A. Elbatsh, and E. Macías, "Gradient rssi filter and predictor for wireless network algorithms and protocols," *Network Protocols and Algorithms*, vol. 2, 2010, pp. 1–26.
- [5] F. Subhan, S. Ahmed, K. Ashraf, and M. Imran, "Extended Gradient RSSI Predictor and Filter for Signal Prediction and Filtering in Communication Holes," *Wireless Personal Communications*, 2015, pp. 1–18.

All-Path Bridging: Path-exploration VS Path-computation on data centers and campus networks

Elisa Rojas, Guillermo Ibañez, Jose Manuel Gimenez-Guzman, Juan A. Carral,

Alberto Garcia-Martinez, Isaias Martinez-Yelmo, Jose Manuel Arco .

Departamento de Automática,
Universidad de Alcalá
Escuela Politécnica Superior, N-II Km 32. 28871 Alcalá de Henares, Madrid, España.

Departamento de Ingeniería Telemática,
Universidad Carlos III de Madrid
Avenida de la Universidad 30. 28911 Leganés, Madrid, España

elisa.rojas@uah.es, guillermo.ibanez@uah.es, josem.gimenez@uah.es, juanantonio.carral@uah.es,
alberto@it.uc3m.es, isaias.martinezy@uah.es, josem.arco@uah.es .

Resumen- Hoy en día, los protocolos de estado de enlaces que calculan múltiples caminos más cortos predominan en redes de centros de datos y redes campus. Estos protocolos pueden funcionar bien en la capa de enlace (TRILL, SPB) o en la capa de red (OSPF). No obstante, estas propuestas no se adaptan totalmente a las variaciones del tráfico en tiempo real y normalmente requieren de algoritmos bastantes complejos para realizar el balanceo de carga de manera eficiente.

Para paliar estas deficiencias, en este trabajo, presentamos All-Path que consiste en una evolución de los puentes transparentes tradicionales que permite liberarlos de las restricciones del protocolo Spanning Tree (STP). All-Path se basa en la exploración simultánea de múltiples caminos mediante broadcast para realizar el balanceo de carga en lugar del cómputo de los mismos basado en la obtención de la topología de la red mediante los protocolos de estado de enlace. Además, sustituye al clásico mecanismo de prevención bucles mediante STP por una sencilla modificación del aprendizaje de direcciones MAC con un temporizador de bloqueo. En concreto, se va a profundizar en el protocolo ARP-Path el cual reutiliza las peticiones estándar ARP de los hosts como mecanismo para explorar y aprender los caminos según se va haciendo necesario en función de los flujos demandados por los host a la red donde están conectados.

Para ver el rendimiento de la solución propuesta, se realiza una comparación del protocolo ARP-Path respecto a SPB en términos de latencia y distribución de carga pudiéndose comprobar que la capacidad de balanceo de carga de nuestra propuesta es ligeramente superior pero mejorando claramente el retardo experimentado por los flujos de datos en la red. Estos resultados se han verificado tanto en simulador de flujos como con un simulador de paquetes para poder controlar tanto efectos macroscópicos como microscópicos y poder hacer estudios escalables en ejecución en función de las topologías seleccionadas para las pruebas.

Como resumen, All-Path ofrece una solución alternativa simple, escalable y con excelente prestaciones respecto a los algoritmos propuestos por SPB o TRILL.

Además de este trabajo, se presentará la evolución más reciente de esta línea de investigación, que por un lado amplía la propuesta de la familia de protocolos All-Path con el protocolo TCP-Path para establecer caminos específicos para flujos TCP concretos. Además, se introducirá brevemente la integración de nuestra solución con switches OpenFlow (switches AOSS) que permite por un lado aumentar la escalabilidad de las redes OpenFlow de una manera fácil y sencilla aumentando su autonomía en labores básicas de conmutación y por otro lado incorpora mecanismos novedosos de recuperación frente a fallos que aprovechan el control centralizado que brinda la arquitectura OpenFlow y la efectividad de la exploración de caminos con broadcast para restablecer la conectividad.

AGRADECIMIENTOS

Este trabajo de investigación ha sido parcialmente soportado por el gobierno de la Comunidad Autónoma de Madrid a través de la financiación del proyecto de investigación TIGRE5-CM (S2013/ICE-2919)

Diseño de una arquitectura para la red de acceso para 5G basada en SDN

Pablo Ameigeiras¹, Juan J. Ramos-Muñoz¹, Laurent Schumacher², Jonathan Prados-Garzon¹,
Jorge Navarro-Ortiz¹, Juan M. Lopez-Soler¹

¹ Dpt. of Signal Theory, Telematics, and Communications, ETSI Informática y de Telecomunicación,
Universidad de Granada (Spain)

{pameigeiras, jjramos, jpg, jorgenavarro, juanma}@ugr.es

² University of Namur

laurent.schumacher@unamur.be

Resumen- El crecimiento exponencial del tráfico de datos y del número de dispositivos conectados, así como la reducción de la latencia y los costes, son considerados como los grandes desafíos de las futuras redes de comunicaciones móviles. Esto motiva que se plantee la necesidad de rediseñar la arquitectura de estas redes. Por ello, en este artículo se propone una arquitectura jerárquica basada en SDN para la red troncal de las redes 5G. En concreto, este trabajo se centra en el diseño de la *nube de acceso* con los objetivos de proporcionar una baja latencia y una escalabilidad similar a la de las redes Ethernet para terminales y dispositivos MTC, incluyendo la gestión de movilidad. La arquitectura de nivel de enlace propuesta reenvía tramas desde y hacia los elementos de red en la frontera (puntos de acceso (AP) y routers) con una etiqueta que identifica al AP al que el terminal está conectado. Un controlador SDN local realiza un seguimiento de la localización de los usuarios y actualiza esta información en los elementos de red en la frontera. Además, se propone delegar en controladores SDN locales la gestión de las operaciones no escalables, tales como el envío de mensajes de difusión, y de procedimientos de gestión de red.

Palabras Clave- 5G, SDN, access cloud.

RESUMEN EXTENDIDO

En este artículo se propone una arquitectura para la red troncal de 3 niveles jerárquicos, denominados *nube de acceso* (AC), *nube distribuida regional* (RC) y *nube centralizada nacional* (NC).

En esta propuesta, cada terminal tiene como identificador único extendido (EUI64) su dirección MAC. La *nube de acceso* (AC) proporciona una comunicación similar a Ethernet, utilizando para ello switches OpenFlow que conectan los puntos de acceso (APs). El tamaño de una AC viene determinado por los límites de la tecnología de nivel de enlace, si bien se puede subdividir una vez que éste se ha alcanzado. Además, se prevén atajos (*local breakouts*) para reducir la carga de la red troncal y reducir la latencia.

La arquitectura incluye un plano SDN que se encarga de gestionar y orquestar las funciones del plano de control. El controlador SDN se compone de *puntos de decisión de políticas* (PDPs) distribuidos por la arquitectura y tan cercanos como sea posible de los dispositivos que controlan. En la AC, los APs y los switches OpenFlow son *puntos de ejecución de políticas* (PEPs) controlados por PDPs que forman un *controlador SDN local* (L-SDN). En la RC, se encuentran routers de capa 3 que son gestionados por un *controlador SDN regional* (R-SDN). De forma similar, existe un *controlador SDN centralizado* (C-SDN) en la NC. El

objetivo de esta división es ejecutar parcial o totalmente las funciones del plano de control de forma local, mejorando así el retardo y la escalabilidad.

Arquitectura y procedimientos de la nube local

En este trabajo proponemos implementar una nube de acceso 5G basada en una arquitectura SDN de nivel de enlace. En el plano de datos, la AC tiene como elementos de red de frontera (ENEs) APs y un router de acceso que la conecta al resto de la red troncal 5G. Estos ENEs se interconectan a través de switches OpenFlow, utilizando etiquetas para su encaminamiento. En el plano de control, un controlador L-SDN gestiona todos los elementos de red. La Fig. 1 resume esta arquitectura.

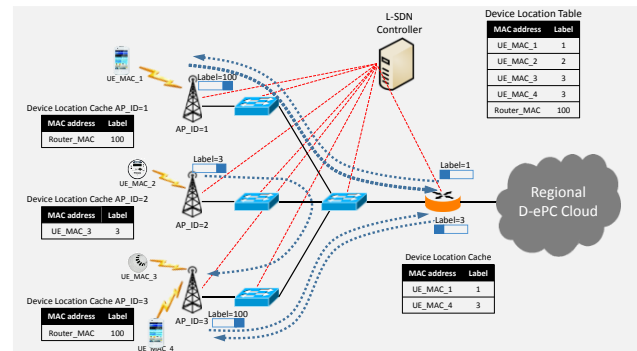


Fig. 1. Arquitectura de red de acceso para 5G.

Los principales procedimientos diseñados en este trabajo son la autoconfiguración de los terminales y la movilidad. La autoconfiguración de los terminales se basa en el uso de mecanismos basados en IPv6 en los que se eliminan los mensajes por inundación gracias al controlador L-SDN. La movilidad entre APs se reduce a unos mensajes de señalización que provocan, gracias al controlador L-SDN, un cambio en las tablas de flujos de los switches, permitiéndose tanto *seamless handovers* como *lossless handovers*.

Un primer análisis permite concluir que la arquitectura y procedimientos propuestos son eficientes, trayendo además otros beneficios propios de las redes SDN, como son la reducción de costes, la flexibilidad, la baja latencia y la alta escalabilidad.

Design of a Synchronous Ethernet in a SDN architecture

Raúl Suárez, David Rincón, Sebastià Sallent

Network Engineering Department
 Universitat Politècnica de Catalunya
 Castelldefels, 08860 Barcelona, Spain

raul.suarez.marin@estudiant.upc.edu, drincon@entel.upc.edu, sallent@entel.upc.edu

Summary- There is an increasing need of synchronization and clock frequency distribution in core and access networks. Synchronous Ethernet (SyncE) is a technology that is able to distribute clock frequency over Ethernet networks. Software-Defined Networking (SDN) is changing the landscape of network management by separating the control and data planes. OpenFlow is currently the most popular southbound interface between the control and data planes in SDN networks. In order to operate SyncE networks under the new SDN paradigm and exploit its advantages, an extension of OpenFlow messages and operations is needed. This paper discusses how SDN and SyncE can interoperate, proposes a set of OpenFlow extensions that are backwards compatible with SyncE equipment, and presents results from a an implementation.

Keywords- SyncE, SDN, OpenFlow, 5G networks, Synchronous OpenFlow, OpenDaylight, Open Vswitch

I. INTRODUCTION

Telecommunications operators are migrating from circuit-based, to packet-switching technologies, due to the inherent advantages of the latter approach. This is the case of 4G/LTE/5G networks. This raises a question: is there any technological solution able to integrate both packet switching and synchronization distribution capabilities?

SyncE is such a technology: the well-known, cheap, scalable Ethernet data plane, with the addition of special messages that convey synchronization information. Regarding timing, SyncE nodes are similar to SDH/SONET nodes: the reference clock is obtained from the signal received from a specific input port and it is used to correct the local clock [1]. Nodes exchange Synchronization Status Messages (SSM) in order to identify the quality of the clocks, and thus deciding (in a distributed way) the best topology for the tree-like clock chain.

SDN separates the control and data planes and introduces a centralized controller, an external centralized entity that manages the network configuration and forwarding functions based on policies defined by the network operator. The controller communicates with simple, cheap switching nodes through standardized interfaces, being OpenFlow (OF) the most popular [2].

II. SDN ENABLED SYNC E ARCHITECTURE

The proposed architecture implements SyncE in SDN networks that allows network operators to introduce synchronization in new SDN network deployments and it eases network management in SyncE networks. Having a centralized entity that manages the network makes that the

configuration of every device is located in the controller, and allows a more efficient and faster management and configuration of the synchronization tree.

These requirements can be programmed in the application that controls the synchronization in the network.

The Openflow SDN-enabled SyncE network architecture is shown in Fig. 1.

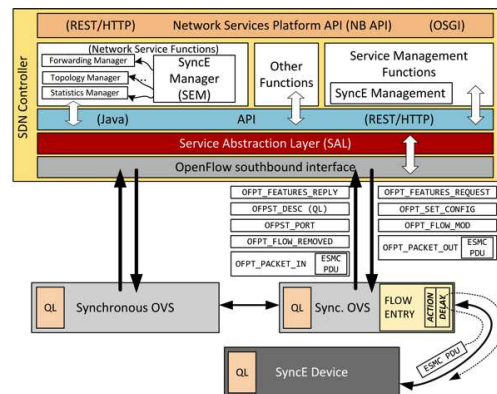


Fig. 1 Proposed architecture

III. IMPLEMENTATION AND TESTING

An implementation of the OF extensions for SyncE enabled SDN networks has been developed in a scenario that involves Open vSwitch as NEs in the data plane and an OpenDaylight SDN controller. The application hosted in OpenDaylight is able to manage SyncE-related parameters and use them in order to create a synchronization tree.

IV. CONCLUSIONS

This work has achieved frequency synchronization using a SDN-enabled Synchronous Ethernet architecture solution by means of OpenFlow extensions.

ACKNOWLEDGMENTS

This work has been supported by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2013-47960-C4-1-P.

REFERENCES

- [1] J. Aweya, "Emerging Applications of Synchronous Ethernet in Telecommunication Systems", IEEE Circuits and Systems Magazine, vol.12, no.2, pp. 56-72, 2012.
- [2] Open Networking Foundation. OpenFlow specifications. Available at <https://www.opennetworking.org/sdn-resources/onf-specifications/openflow>

Emerging Real-time Services: Optimizing Traffic by Smart Cooperation in the Network

Julián Fernández Navajas, José Ruiz Mas, José María Saldaña Medina.
 Grupo CeNITEQ– Instituto de Investigación en Ingeniería de Aragón
 Dpt. IEC. EINA, Universidad de Zaragoza
 Edif. Ada Byron, 50018, Zaragoza
 {navajas, jruiz, jsaldana }@unizar.es

Dan Wing, Muthu A. M. Perumal, Michael Ramalho, Cisco Systems
 Gonzalo Camarillo, Ericsson Research

Fernando Pascual, Diego R. Lopez, Manuel Núñez, David Flórez, Juan A. Castell, Telefonica I+D
 Tomaso de Cola, Matteo Berioli, DLR (German Aerospace Center)

Resumen

El diseño inicial de Internet no contemplaba la posibilidad de entregar la información cumpliendo requisitos temporales. Así, la red se diseñó con el objetivo de enviar paquetes grandes, donde cada unidad de información debe disponer de una serie de cabeceras de distintos niveles para poder llegar a su destino. Cada tecnología tiene un tamaño máximo de paquete, y si se envían paquetes de ese tamaño o algo inferiores, el coste de las cabeceras (*overhead*) es asumible, al repartirse entre un gran número de bytes de información. Sin embargo, si se desea utilizar Internet para soportar servicios en tiempo real, es necesario enviar actualizaciones frecuentes de información y esto obliga a que habitualmente los paquetes sean pequeños, pues los requerimientos temporales impiden esperar con el fin de acumular más bytes de información. En esta situación, los operadores de red están observando un cambio en la distribución agregada del tráfico que deben transportar sus redes y la proporción de flujos de paquetes pequeños aumenta [1].

Vemos, por tanto, que surgen nuevos servicios que generan flujos extremo a extremo de larga duración, con baja eficiencia en el transporte y que comparten un mismo camino de red con otros flujos. Así pues, la optimización de estos flujos mejoraría la eficiencia de la red, consiguiendo ahorros efectivos de ancho de banda y la disminución del número de paquetes por segundo (pps) que la red debe gestionar.

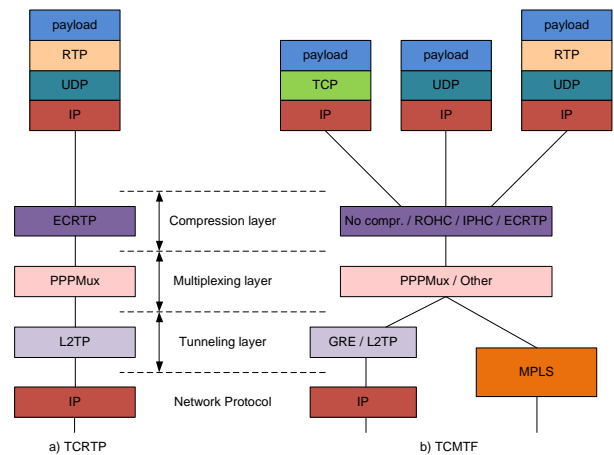
Para reducir el *overhead* se han desarrollado algoritmos de compresión de cabeceras (RFC 3545, RFC 2507, RFC 5225), que aprovechan la redundancia de muchos de los campos, que son iguales en todos los paquetes de un flujo. El problema de estos algoritmos es que sólo pueden funcionar punto a punto, puesto que un paquete con una cabecera incompleta no puede viajar por Internet. Para enviarlos extremo a extremo debe usarse un protocolo de tunelado (p. ej. RFC 2661), que vuelve a añadir gran parte del *overhead* que queríamos evitar. Pero si la compresión de cabeceras se combina con la multiplexión (RFC 3153) de varios paquetes dentro de uno más grande, entonces el *overhead* causado por el túnel se reparte, y pueden conseguirse ahorros de ancho de banda mucho mayores.

Para conseguir la optimización del tráfico se propone la combinación de un conjunto de métodos: compresión de cabeceras, multiplexión y tunelado (TCM). Sobre dichos

métodos se ha realizado el análisis correspondiente, que consta de la propuesta de aplicación de los métodos, y de un estudio analítico de eficiencia, retardo, pérdidas y calidad percibida por los usuarios de los servicios.

Por lo tanto vemos que una parte fundamental del trabajo es la definición de dicha propuesta y su pila de protocolos. Esta nueva propuesta no parte de cero sino que se basa en la RFC 4170 (uno de sus autores forma parte del presente trabajo) que proporciona la optimización para RTP (TCRTP, Tunnelerd multiplexed Compressed RTP). Sin embargo la nueva propuesta no se limita a RTP sino que puede aplicarse directamente a los protocolos UDP y TCP.

En la siguiente figura se observa la propuesta realizada (TCMTF, Tunneling Compressed Multiplexed Traffic Flows) comparándola con la ya existente (TCRTP).



Protocol stack of a) TCRTP and b) TCMTF

Para finalizar, se han presentado una serie de gráficas sobre el análisis realizado en cuanto a la mejora de la eficiencia y el ahorro de paquetes por segundo, sobre tres servicios en concreto: flujos de VoIP G.729, Counter Strike 1 y World of Warcraft, cuyo detalle podemos ver en el trabajo original: J. M. Saldaña; et al. 2013. *Emerging Real-time Services: Optimizing Traffic by Smart Cooperation in the Network*. IEEE Communications Magazine. 51-11, pp.127-136. ISSN 0163-6804. Estos servicios han sido escogido porque cumplen los requisitos de generar tráfico de tramas pequeñas en tiempo real y soportan un análisis sencillo de percepción de usuario.

Recuperación de Información desde diferentes perspectivas

Diego Fernández, Víctor Carneiro, Francisco Nóvoa de Manuel, Xacobe Macía da Silva, Fidel Cacheda
Grupo de Ingeniería Telemática. Universidad de A Coruña.
Campus de Elviña s/n, 15071, A Coruña.
{diego.fernandez, victor.carneiro, fjnovo, xacobe.macia, fidel.cacheda}@udc.es

Algunas de las principales líneas de trabajo del grupo de Ingeniería Telemática de A Coruña son: Filtrado Colaborativo (CF), Sistemas de Recuperación de Información (IR) Distribuidos y Motores de Búsqueda Web. A continuación se recogen los avances más importantes del grupo en estas líneas.

A. Filtrado Colaborativo

Las técnicas de CF se emplean para la recomendación de productos y son especialmente populares en comercio electrónico, tanto para la sugerencia de listas de productos como para la predicción de la valoración de nuevos productos. Estas técnicas, las más empleadas y maduras, trabajan sólo con puntuaciones de usuarios sobre productos. En [1] el grupo realizó un estudio del estado del arte de estos algoritmos y propuso un nuevo algoritmo basado en tendencias.

Con el paso del tiempo la calidad de las recomendaciones ha ido mejorando. A pesar de ello, la mayoría de las técnicas de CF presentan problemas en lo que se refiere a eficiencia y escalabilidad. Entre las existentes, los algoritmos *k-Nearest Neighbours* (kNN) son muy populares. Sin embargo, poseen ciertos inconvenientes en su rendimiento, principalmente en la fase de selección de vecinos. Con frecuencia, dichos inconvenientes se mitigan pre-computando los vecinos. En las aplicaciones reales esto no es suficiente, y es necesario mejorar también la fase de recomendación.

Muchos sistemas están apostando por realizar un procesado distribuido de datos gracias, en parte, a las tecnologías de *Big Data*. Estas tecnologías se pueden emplear para implementar sistemas de CF y reducir el tiempo necesario para hacer las recomendaciones. En [2] el grupo propuso una arquitectura distribuida para la recomendación de productos empleando algoritmos kNN.

B. Sistemas de IR distribuidos

El incremento de información en la web hace necesario el empleo de sistemas de IR distribuidos. Éstos constan de *brokers*, que interactúan con los usuarios, y de servidores de consultas, que procesan las consultas y generan una lista de resultados.

Para probar sistemas de IR distribuidos para su uso a gran escala, se puede acudir a sistemas de virtualización. En los experimentos realizados por el grupo, KVM alcanzó los mejores resultados. Además, se obtuvo una fuerte correlación lineal entre los tiempos de distribución en KVM y los reales.

El grupo también ha trabajado en técnicas de *Query Scheduling* para escoger la mejor réplica de cada servidor reduciendo el tiempo de respuesta a una consulta en un sistema de IR distribuido y replicado.

Otro aspecto que se ha considerado es el denominado *Green IR*, en donde existe poca investigación, a pesar de que los motores de búsqueda comerciales consumen gran cantidad de energía. En el grupo se ha definido el número de máquinas necesarias a lo largo del día, conforme varía el número de consultas. Así, usando un modelo matemático, se reduce el consumo de energía de un motor de búsqueda sin degradar la eficiencia global del sistema [3]. Originando menos de un 1% de incremento en la latencia por consulta, es posible reducir el consumo de energía hasta en un 68%.

C. Motores de búsqueda web

Para poder llevar a cabo búsquedas sobre grandes cantidades de datos, los motores de búsqueda web disponen de *crawlers*, que permiten recorrer la web con el objetivo de crear un repositorio de recursos sobre los que poder realizar dichas búsquedas. Sin embargo, existen diversos desafíos, en los que el grupo ha trabajado: el tratamiento de la web oculta de lado cliente y servidor, la detección de páginas basura [4] y la actualización de contenidos [5].

En primer lugar, se realizaron diversos estudios para analizar la evolución de la Web, y estudiar y caracterizar la web oculta, y cómo los *crawlers* la tratan.

Los métodos existentes para la detección de páginas basura presentaban muchos problemas tanto de eficiencia como de eficacia. Para mitigar estos problemas en el grupo se desarrollaron nuevas técnicas basadas en contenido. De este modo el *crawler* no dedica recursos innecesariamente a gestionar este tipo de páginas, mejorando así la calidad de sus repositorios

Finalmente, se propuso una arquitectura para que el *crawler* procese únicamente aquellas páginas que hayan cambiado, incrementando así su eficiencia.

REFERENCIAS

- [1] Fidel Cacheda, Víctor Carneiro, Diego Fernández, Vreixo Formoso. "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems". 5:2:1-2:33, Febrero 2011.
- [2] Vreixo Formoso, Diego Fernández, Fidel Cacheda, Víctor Carneiro. "Distributed architecture for k-Nearest Neighbors recommender systems". World Wide Web. Internet and Web Information Systems. Volume 18, Issue 4, pp 997-1017, 2014. Springer.
- [3] Ana Freire, Craig Macdonald, Nicola Tonellotto, Iadh Ounis and Fidel Cacheda. "A self-adapting latency-Power Trade-off Model for Replicated Search Engines". WSDM 2014.
- [4] Víctor M. Prieto, Manuel Álvarez, Fidel Cacheda. "Soft-404 Pages, a Crawling Problem". Journal of Digital Information Management (JDIM). Vol. 12, issue 2, pp. 73-92, 2014.
- [5] Víctor M. Prieto, Manuel Álvarez, Víctor Carneiro, Fidel Cacheda. "Distributed and Collaborative Web Change Detection System". Computer Science and Information Systems Journal (ComSIS). ComSIS Consortium. Vol. 12, issue 1, pp. 91-114, 2015.

Monitorization Estadística Multivariante para Ciberseguridad: Sistema VERITAS

José Camacho, Gabriel Maciá Fernández, Pedro García Teodoro
Departamento de Teoría de la Señal, Telemática y Comunicaciones,
Universidad de Granada

C/ Periodista Daniel Saucedo Aranda s/n 18071 GRANADA
josecamacho@ugr.es, gmacia@ugr.es, pgteodor@ugr.es

Roberto Therón

Departamento de Informática y Automática
Universidad de Salamanca,
Plaza de los Caídos s/n, 37008 SALAMANCA
theron@usal.es

Resumen- Derivado del número de vulnerabilidades e incidentes reportados de forma creciente en las últimas fechas, la seguridad en redes y sistemas constituye uno de los tópicos más relevantes en la actualidad en relación a las TIC. Dadas las cada vez mayores prestaciones, capacidades y complejidad de los sistemas y servicios que sustentan la intercomunicación global, la gestión de la seguridad se ha convertido en un problema Big Data; haciéndose precisa la disposición de grandes cantidades de datos a partir de las cuales poder concluir la ocurrencia (o no) de situaciones ilegítimas y/o indeseadas. En este marco, la monitorización estadística multivariante de redes (MSNM, del inglés Multivariate Statistical Network Monitoring) es una propuesta novedosa de detección de anomalías de sistemas en red con aplicación a la seguridad. MSNM hereda la metodología de detección desarrollada durante medio siglo en el campo de la monitorización industrial, basándose en el uso de la técnica de análisis por componentes principales (o PCA, del inglés Principal Component Analysis) y estadísticos de monitorización multivariante. En esta presentación se introduce la problemática del diseño de sistemas MSNM, poniendo el acento en sus ventajas y retos de cara a su aplicación a la ciberseguridad. En relación directa con ello se presenta el proyecto nacional de investigación VERITAS (<http://nesg.ugr.es/veritas>), el cual tiene por objeto la extensión de la metodología MSNM haciendo uso de técnicas jerárquicas, para la reducción del tráfico asociado a parámetros de seguridad, y la mejora de la privacidad en modelos de negocio, donde la seguridad es externalizada.

Palabras Clave- ciberseguridad, detección de anomalías, análisis multivariante

Este trabajo está financiado por el Ministerio de Economía y Competitividad, con fondos FEDER, a través del proyecto TIN2014-60346-R.

XII Jornadas de Ingeniería de Telemática (JITEL)

Joint Strategy for LTE Resource Allocation: Multicast Subgrouping & Unicast Transmissions

Alejandro de la Fuente, Raquel Perez Leal, Ana Garcia Armada
Department of Signal Theory and Communications
University Carlos III of Madrid
Email: afuente@tsc.uc3m.es, agarcia@tsc.uc3m.es, rpleal@tsc.uc3m.es

Abstract—Mobile broadband services are growing rapidly in the last few years due to the deployment of Long Term Evolution (LTE) cellular networks. Among them, multicast services can be provided using Evolved Multimedia Broadcast and Multicast Service (eMBMS), available with 3rd Generation Partnership Project (3GPP) release 9, which can deliver broadcast/multicast content using a single-frequency network mode. This means sending the same multimedia content to a mass audience within a specific area. The utilization of the Conventional Multicast Scheme (CMS) approach for multicast resource allocation presents intrinsic inefficiencies, because of the different channel conditions of the users which demand the service. This paper proposes a Joint Multicast Subgrouping and Unicast Transmissions (JMSUT) strategy for resource allocation, which consists of the use of the multicast and the unicast transmissions, by means of the subframes reserved by the LTE standard for each purpose, to deliver a multicast service. The goal of the JMSUT algorithm is to maximize the service throughput whereas it guarantees the fulfillment of the Quality of Service (QoS) requirements of every user. This paper solves the former maximization problem of the joint resource allocation; on the one hand, splitting the multicast resources into different subgroups that transmit the same content with different Modulation and Coding Scheme (MCS), and on the other hand, the users with worst channel conditions are served by means of the unicast transmissions.

I. INTRODUCTION

Mobile data traffic is growing rapidly in the last few years and this growth is expected to become bigger in the upcoming years, especially in multimedia services. Of course, the growing demand of multimedia services in mobile networks poses new challenges in the way these services can be provided. New techniques must be developed to guarantee the scalability for large amount of users, since in the near future more devices will be connected leading to what is known as "everything connected". In 2020, approximately 25 billion interconnected devices are expected [1].

Broadcasting and multicasting are expected to be promising enablers of an easy access to the ubiquitous multimedia experience through mobile terminals [2]. An evolved architecture is required to support Evolved Multimedia Broadcast and Multicast Service (eMBMS) transmission in Long Term Evolution (LTE) network. Such an architecture is detailed in the 3rd Generation Partnership Project (3GPP) specifications [3], there are new logical network entities proposed for eMBMS operation, which enable a point-to-multipoint service that allows data transmissions from a single source to multiple recipients. Consequently, the scalability of broadcast

and multicast transmissions in mobile networks is improved. Furthermore, Multicast/Broadcast over Single Frequency Network (MBSFN) has been proposed as an enhancement of eMBMS [3], avoiding the destructive interferences in the areas where the coverage overlaps, and maintaining the performance that would otherwise gradually decrease as User Equipment (UE) moves away from the base station.

While using multicast transmissions improves the efficient utilization of network resources, it requires setting equal transmission parameters to all the users in the MBSFN area. Consequently, in multicast transmissions, the Modulation and Coding Scheme (MCS) is unique and set by upper layers. Therefore, the multicast transmission throughput in the MBSFN area is established by the MCS and the transmission bandwidth [4]. Differently, unicast transmissions can use link adaptation and channel dependent scheduling, based on the Channel Quality Indicator (CQI) the user sends periodically to the Evolved Node B (eNodeB). Therefore, Evolved Universal Terrestrial Radio Access (E-UTRA) can dynamically allocate resources, both the number of Physical Resource Block (PRB) and MCS, to the UEs at each Transmission Time Interval (TTI) [4].

In multicast transmissions, the resource allocation using the Conventional Multicast Scheme (CMS) [5] is based on a conservative approach, where the data rate is restricted by the user that presents worst channel conditions. Of course, this approach maximizes the fairness among multicast users, however the throughput performance in the multicast area is highly inefficient and users with good channel conditions do not achieve as high as possible bit rates.

Recent research, motivated by these issues, studies solutions to improve the multicast service throughput, at the time fairness among users is achieved. In [6], the basic idea of splitting any multicast group into subgroups is utilized and, after that, the strategy applies Adaptive Modulation and Coding (AMC) schemes, which enables a more efficient exploitation of multi-user diversity. This proposal distributes the users into subgroups by solving an optimization problem to improve the service throughput while guaranteeing fairness among multicast members.

In [7], a Joint Multicast/Unicast Scheduling (JMUS) was proposed to maximize the overall throughput in the MBSFN area. The proposed technique combines unicast and multicast transmissions to guarantee a target bit rate for all the users

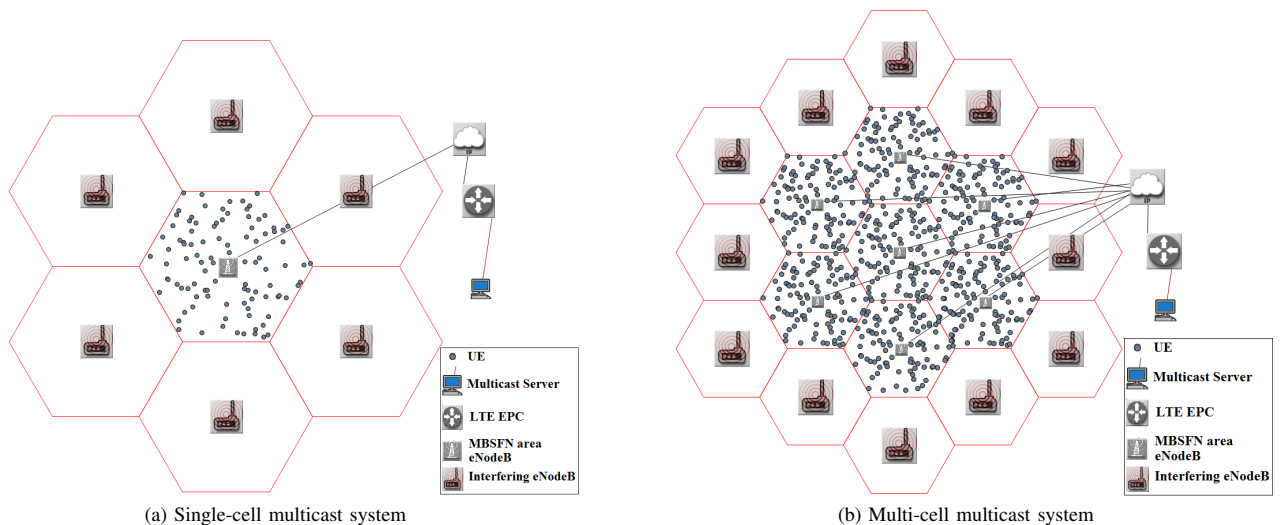


Fig. 1. System model reference scenario

demanding a multicast service. The optimal MCS used for multicast transmission are obtained each frame. The JMUS with dynamic optimization enhances the Quality of Service (QoS) performance compared to pure unicast, pure multicast, or scheduling techniques without dynamic optimization.

In this paper, we propose a new Joint Multicast Subgrouping and Unicast Transmissions (JMSUT) strategy for resource allocation in an LTE multicast service delivery. This approach combines the unicast transmissions using the LTE subframes reserved for that purpose, and the multicast transmissions using the remaining subframes. The goal of the JMSUT algorithm is to maximize the service throughput whereas it guarantees the fulfillment of the QoS requirements of every user. This paper solves the former maximization problem of the joint resource allocation. On the one hand, the total resources for multicast transmissions are splitted into different subgroups that transmit the same content with different MCS, according to the solution of the joint optimization problem. On the other hand, the users with worst channel conditions are served using a scheduling metric for guaranteed data-rate proposed in [8] to allocate the resources reserved for unicast transmissions. The results presented in this work show that the joint use of multicast subgrouping and unicast transmissions allows important enhancements in the multicast service throughput, at the time the QoS requirements are guaranteed for all the users.

By multicast service we refer to a streaming or downloading service delivered to all the users in the system, while we denote by multicast transmission the utilization of the Physical Multicast Channel (PMCH) by the eNodeB to send the same data to all the users, and by unicast transmission the use of Physical Downlink Shared Channel (PDSCH) by the eNodeB to send the data to each UE [4]. That is, the multicast service can be delivered to a given user by either multicast (shared)

or unicast (dedicated) transmission.

The rest of the paper is organized as follows. In Section II, the system model used is described. The proposed JM-SUT strategy is detailed in Section III. The results of the performance evaluation are presented in Section IV. Finally, in Section V, the conclusions and future work are explained.

II. SYSTEM MODEL

This paper utilizes two reference scenarios deployed to evaluate the resource allocation strategies used to deliver a multicast service in LTE systems.

Fig. 1(a) depicts a single-cell multicast system where an LTE eNodeB provides a multicast service to UEs which are uniformly distributed within the cell. Around this single-cell, a tier of 6 interfering eNodeBs is deployed.

Fig. 1(b) illustrates a multi-cell multicast system, where 7 LTE eNodeBs are coordinated in a single MBSFN. The multicast service is provided in the 7-cell area to UEs which are uniformly distributed within each cell. Around the 7-cell MBSFN area, we consider one tier of 12 eNodeBs operating on the same frequency and transmission power as the 7 eNodeBs in the MBSFN area.

In LTE systems, radio resources are allocated into the time/frequency domain [10]. In the time domain, they are distributed every TTI of 1 ms. Time is organized in frames, each one composed of 10 consecutive TTIs or subframes. In addition, each TTI is made of two time slots with 0.5 ms length. Each time slot corresponds to 7 Orthogonal Frequency Division Multiplexing (OFDM) symbols with normal cyclic prefix (default configuration for unicast transmissions), or 6 OFDM symbols with extended cyclic prefix (recommended for MBSFN configuration for multicast transmissions). In the frequency domain, the total bandwidth is divided in sub-channels of 180 kHz, each one with 12 consecutive 15 kHz

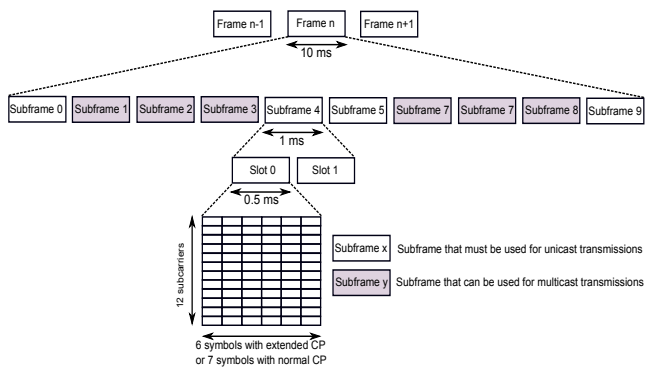


Fig. 2. LTE-FDD frame/subframe structure

OFDM sub-carriers. The LTE frame/subframe structure for Frequency Division Duplexing (FDD) is illustrated in Fig. 2.

A PRB is the smallest radio resource unit that consists of a 2D radio resource, over two time slots in the time domain, and one sub-channel in the frequency domain. As the sub-channel size is fixed, the number of PRBs varies according to the system bandwidth configuration (e.g. 15 PRBs for system bandwidth of 3 MHz). The PRBs are managed by the packet scheduler in the eNodeB in three phases. Firstly, the information on the schedulable services is collected based on the buffer states. Secondly, in the time domain, the UEs to serve are selected and their QoS constraints are established. Finally, in the frequency domain, the Radio Resource Management (RRM) procedures are executed to meet the QoS constraints.

To implement a channel-aware resource allocation strategy, UEs' CQI are assumed to be known at the eNodeB [5]. CQI is estimated at each UE from the Signal to Interference plus Noise Ratio (SINR) measurement of its radio channel, and sent to the eNodeB using feedback. The eNodeB utilizes this information forwarded by the UEs every CQI Feedback Cycle (CFC) to allocate the resources, determining the MCS used for each unicast transmission.

Furthermore, in multicast mode, the Physical Uplink Control Channel (PUCCH) is used in the uplink to send control messages, e.g. channel state information, from the multicast members to the eNodeB. CQI gives information about the maximum MCS that a UE can decode [4]. LTE systems utilize 15 CQI levels that correspond to the SINR measured by the terminal, and at the same time, these CQIs are related to the maximum MCS supported by the terminal for a correct decoding (refer to Table I). Thus, a UE reporting a determined CQI value can successfully demodulate only the service delivered with an MCS whose index is equal or lower than the CQI index reported.

eMBMS in LTE systems standardizes the utilization upto 6 subframes of the LTE frame for multicast transmissions using PMCH [3]. The remaining frames (at least 4 subframes) are reserved for unicast transmissions using PDSCH. According to eMBMS standardized service, the resource allocation strategy proposed here searches the solution that allocates jointly, on

TABLE I
SINR-CQI-MCS MAPPING [9]

SINR (dB)	CQI index	Modulation scheme	Code rate	Spectral Efficiency (bit/s/Hz)
-4.63	1	QPSK	0.076	0.1523
-2.6	2	QPSK	0.120	0.2344
-0.12	3	QPSK	0.190	0.3770
2.26	4	QPSK	0.300	0.6016
4.73	5	QPSK	0.440	0.8770
7.53	6	QPSK	0.590	1.1758
8.67	7	16QAM	0.370	1.4766
11.32	8	16QAM	0.480	1.9141
14.24	9	16QAM	0.600	2.4063
15.21	10	64QAM	0.450	2.7305
18.63	11	64QAM	0.550	3.3223
21.32	12	64QAM	0.650	3.9023
23.47	13	64QAM	0.750	4.5234
28.49	14	64QAM	0.850	5.1152
34.6	15	64QAM	0.930	5.5547

the one hand, the PRBs in the multicast subframes splitted into different subgroups that deliver the service using different MCS, and on the other hand, the PRBs in the unicast subframes to serve the UEs with worst channel conditions, so that they can fulfill the QoS requirements.

III. RADIO RESOURCE MANAGEMENT STRATEGY USING JMSUT

A multicast service is delivered in an MBSFN area using a dedicated LTE bandwidth. An LTE system can use multicast and unicast transmissions to provide the service to all the users. The proposed RRM strategy uses a JMSUT. On the one hand, the RRM algorithm searches the optimal allocation of the resources in the multicast subframes, splitting them into multicast subgroups that deliver the service using different MCS. On the other hand, the RRM uses the unicast QoS-aware scheduling proposed in [8] to deliver the service using unicast transmissions to the UEs with worst channel conditions. Consequently, the JMSUT aims to maximize the service throughput and, at the same time, guarantees the QoS requirements for all the users demanding the service.

The proposed JMSUT strategy can be splitted into different phases that are described in the following subsections.

A. CQI collection

The first step consists of the collection by the eNodeBs of the CQI feedback from the UEs placed in their MBSFN area which are demanding the multicast service. For each CFC, the eNodeB creates a vector C with all the UEs CQI, such as $C = c_1, c_2, \dots, c_n$, where c_i is the CQI reported by user i and n is the number of multicast members served by the eNodeB.

B. Multicast subgroup creation

The proposed JMSUT algorithm splits the multicast member into different multicast subgroups, and the members with

worst channel conditions can be attended using unicast transmissions. Each multicast subgroup delivers the service using different MCS in order to serve the users that support the decoding of this scheme with a Block Error Rate (BLER) less than 10% [4].

C. Joint multicast and unicast resource allocation

Radio resources available to deliver the service depend on the bandwidth reserved for that purpose. These PRBs are allocated in multicast subframes (upto 6 subframes of the LTE frame) and unicast subframes.

The resource allocation algorithm proposed works such as whereas a UE reports a CQI that is equal or greater than the lowest multicast subgroup with PRBs allocated, this user will be served by the multicast subgroup closer to the CQI reported and whose MCS can be decoded by the user. On the other hand, only the users which are reporting CQI lower than the lowest multicast subgroup are served using the unicast transmissions reserved in the LTE frame.

Therefore, this algorithm is based on a service throughput maximization problem. This problem presents several constraints, such as the minimum bit rate per user, or the number of PRBs available to deliver the service. The following section describes the Maximum Throughput (MT) algorithm used for this proposal.

D. Maximum Throughput Optimization Problem

The Maximum Throughput (MT) strategy searches to maximize the throughput that can be achieved, enhancing the capacity results obtained using a CMS approach. This strategy is based on a maximization problem of a function cost, that consists of the sum of the data rate of all the members of the multicast service. Furthermore, a minimum bit rate per user has been established to guarantee QoS requirements. Consequently, this maximization problem can be expressed as follows for the reference scenarios:

$$\underset{R}{\text{maximize}} \quad \sum_{i=1}^n d_{c_i}^R \quad (1)$$

$$\text{subject to} \quad M + U = 10 \quad (1a)$$

$$M \leq 6 \quad (1b)$$

$$d_{c_i}^R \geq b_{min} \quad \forall i \quad (1c)$$

$$\sum_{i=1}^G r_i = K \quad (1d)$$

$$\sum_{i=1}^{n_u} K_i \leq U \times K \quad (1e)$$

where $R = \{r_1, \dots, r_G\}$ is the distribution vector which allocates the PRBs into the different multicast subgroups, $d_{c_i}^R$ denotes the bit rate achieved to deliver the service to user i when the distribution vector R is used to allocate the PRBs among the multicast subgroups, and n is the total number of

TABLE II
SYSTEM PARAMETERS

Parameter	Value
Multi-cell system size	7 eNodeBs
Interference model	1 tier of eNodeBs
eNodeBs geographical overlay	Hexagonal
Inter site distance	500 m
Transmission power	43 dBm
Antenna gain	11.5
Bandwidth	3 MHz
Number of PRBs	15
Downlink base frequency	2110 MHz
Pathloss model	3GPP Urban Macrocell
Multipath channel model	ITU Pedestrian B
eNodeB transmission antennas	1
UEs per eNodeBs	100
UEs distribution	Uniform distribution
Guaranteed bit rate per user UE	20-200 kbps
Pedestrian user speed	3 Km/h
Vehicular user speed	50 Km/h

users demanding the multicast service in the MBSFN area. In (1a), M and U denote the number of subframes reserved by the standard for multicast and unicast transmissions in an LTE frame, respectively. In (1b), the maximum number of subframes that can be reserved for multicast transmissions is 6. In (1c), the minimum bit rate that must be guaranteed for all the users is denoted as b_{min} . In (1d), the maximum number of PRBs to allocate is established, so to that end, G denotes the maximum number of multicast subgroups (in LTE there are 15 different CQI sublevels), and K denotes the number of available PRBs per subframe to deliver the service. In (1e), the total cell capacity is ensured not to be surpassed by the unicast transmissions, so n_u denotes the number of users served by unicast transmissions, and K_i is the amount of resources allocated to user i during the LTE frame.

IV. PERFORMANCE EVALUATION

The performance evaluation has been carried out using the reference scenarios described in the former section for single-cell and multi-cell multicast system, based on LTE standard. These scenarios use 3 MHz bandwidth, thus 15 PRBs are available to deliver the service. A deployment of 100 users, which are multicast members of the service, has been used with a uniform distribution in each cell. Different combinations of users have been evaluated, using static, pedestrian (mobile users at 3 Km/h) and vehicular (50 Km/h) UEs. The JMSUT algorithm has been used with different QoS constraints, guaranteeing a bit rate between 20 and 200 kbps for all the users. It is worth noting that this performance evaluation has been carried out using a dedicated bandwidth of 3 MHz, therefore higher data rates can be achieved for a multicast service delivered to 100 users reserving more bandwidth for this service. Main simulation parameters are listed in Table II.

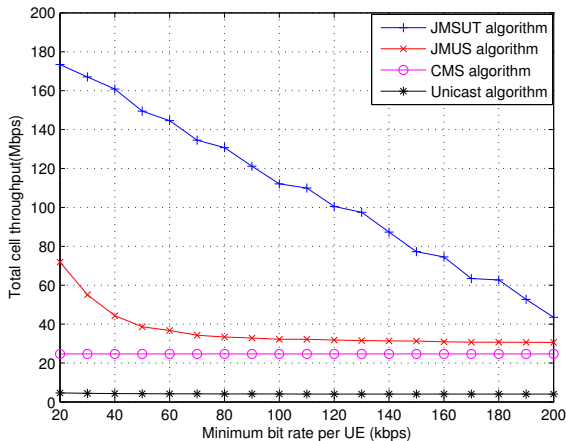


Fig. 3. Performance evaluation using a scenario based on single-cell eMBMS using static and pedestrian users

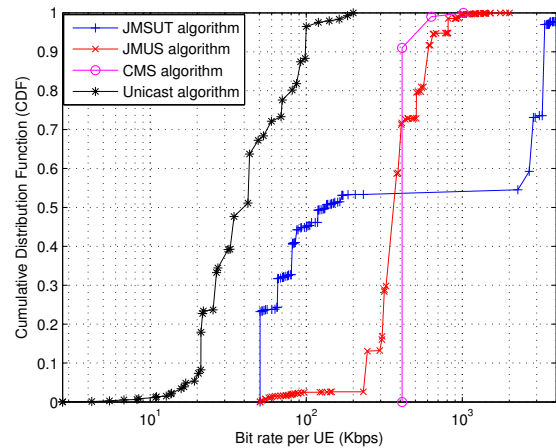


Fig. 4. CDF of users' bit rate in a scenario based on single-cell eMBMS using static and pedestrian users with GBR=50 kbps

This paper shows three different evaluation of the JMSUT strategy use.

The first evaluation consists of a comparison of different resource allocation strategies in a single-cell multicast scenario. The results achieved using the proposed JMSUT strategy are compared with the JMUS proposed in [7], with CMS and with the use of only unicast transmissions.

The second evaluation is based on the multi-cell multicast scenario. A comparison among the results achieved in the central cell and a peripheral cell using both JMSUT and JMUS strategies is presented. In addition, these results are compared with the ones achieved in a single-cell multicast deployment.

The third evaluation compares the results obtained with different combinations of users, such as static, pedestrian and vehicular ones.

A. Comparison of different resource allocation strategies in a single-cell multicast scenario

The first evaluation shows the results achieved using different resource allocation strategies in the single-cell multicast scenario with static and pedestrian users.

On the one hand, Fig. 3 illustrates the total service throughput as a function of the minimum bit rate required for every user. It can be noticed that the use of multicast transmissions highly improves the performance of using only unicast transmissions. Nonetheless, the application of joint resource allocation techniques enhances the throughput results with respect of the most conservative multicast scheduling scheme (CMS). Especially with the use of multicast subgrouping, since we can observe how the JMSUT strategy results in important improvements in service throughput over the use of JMUS strategy. However, as the minimum bit rate per user is increased, this gain in total throughput is decreasing. This is because the resource allocation strategy must ensure that users with worst channel conditions reach this minimum bit rate,

allocating more resources to the groups that are less efficient in terms of throughput.

On the other hand, Fig. 4 shows the Cumulative Distribution Function (CDF) of the bit rate achieved per UE in this scenario, when the minimum bit rate target is established to 50 kbps. It can be noticed that the minimum bit rate required for all the users cannot be guaranteed using only unicast transmissions. On the opposite side, the utilization of CMS guarantees the maximum fairness among all the users, but at the expense of users with good channel conditions are not having benefit of it, and for that reason the total service throughput is low. On the other side, JMSUT not only allows all the users to achieve the minimum required bit rate, but also users that present good channel conditions can obtain higher bit rates, and consequently the total throughput of the service is greatly increased.

It is worth noting that all the strategies evaluated are using all the resources available in the LTE frame, either using only unicast transmissions or using a combination among multicast and unicast,

B. Comparison of different resource allocation strategies in a multi-cell multicast scenario

The second evaluation illustrates the results obtained using the multi-cell multicast scenario with static and pedestrian users.

In Fig. 5, the service throughput achieved in the central cell and a peripheral one, using both JMSUT and JMUS strategies, are presented. Furthermore, these results are compared with the ones achieved using the single-cell multicast scenario. It can be noticed an important throughput gain obtained in multi-cell scenario, especially in the central cell. The use of coordinated transmissions among 7-cells in an MBSFN area improves the channel conditions of the users in the cell edge, especially in central cell. In addition, this improvement in the channel conditions of the users leads to a higher gain using

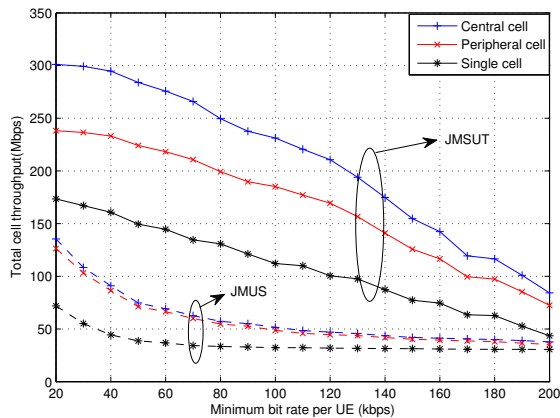


Fig. 5. Performance evaluation of JMSUT and JMUS in single and multi cell scenario using static and pedestrian users

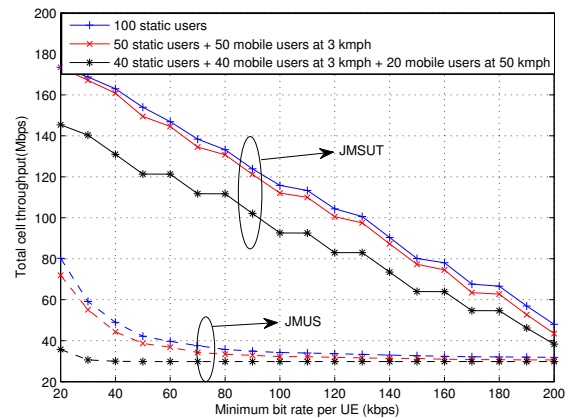


Fig. 6. Performance evaluation of JMSUT and JMUS in a scenario based on single-cell eMBMS and different combinations of static and mobile users

JMSUT instead of JMUS, since the target bit rate required in the users with worst channel conditions can be fulfilled using less amount of resources.

C. Comparison of different resource allocation strategies with different types of users

The third evaluation presents the results achieved in a single-cell multicast scenario using different combinations of users, such as static, pedestrian, and vehicular users.

The service throughput achieved using several deployments of different combinations of users is presented in Fig. 6. It can be observed that scenarios with static and pedestrian users can achieve similar service throughput, slightly better when all the users are static. Only when vehicular users are incorporated to this scenario, the results in the service throughput achieved in the cell suffer an important lowering.

In addition, it is worth noting that the use of JMSUT instead of JMUS results in an important gain in terms of service throughput, regardless the combination of users in the simulated scenario.

V. CONCLUSIONS

This work proposes the use of a joint resource allocation strategy among the unicast and multicast subframes in LTE eMBMS service. It proposes the creation of different multicast subgroups to allocate the available PRBs among them for the multicast subframes, and combine it with the transmissions in unicast subframes used to serve the users with worst channel conditions, in such a way this joint strategy maximizes the service throughput in the cell.

The performance evaluation of the proposed algorithm has shown how it can greatly improve, in terms of service throughput, the results achieved using CMS (i.e. one multicast group based on the user with the worst channel conditions), or JMUS strategy proposed in [7] (i.e. joint multicast and unicast transmissions using only one multicast group based on the joint optimization). At the same time, JMSUT algorithm

allows all the users to achieve the minimum required bit rate. It is worth noting that the creation of different multicast subgroups results in users with good channel conditions can be served using high bit rates, and users with worse conditions can be served with the minimum target bit rate. Thus, the proposed algorithm allows the system to maximize the service throughput in the cell, at the time it guarantees a minimum service level for all the users.

In addition, the evaluation of JMSUT in a multi-cell multicast scenario has presented significant higher performance, in terms of service throughput, than in single-cell multicast scenario. Since the channel conditions of the UEs are enhanced using a 7-cell MBSFN area, especially those users which are placed at cell edge, leading to a higher gain in service throughput when the JMSUT strategy is used.

Finally, the evaluation of scenarios with different combinations of users has illustrated the gain achieved using JMSUT, independently of the mobility of the users. Although the service throughput is lowered with vehicular users, it happens regardless the resource allocation strategy used, and the throughput gain using JMSUT is achieved for all the combinations of users simulated.

In future work, the use of JMSUT strategy with other different cost functions, such as Proportional Fairness (PF) algorithm, will be evaluated in order to maximize the total throughput of the service increasing the fairness among all the users. Furthermore, the introduction of memory to the JMSUT algorithm will be analyzed to increase the fairness among users without lowering the total service throughput.

ACKNOWLEDGMENT

This work was supported in part by the Spanish Ministry of Economy and Competitiveness, National Plan for Scientific Research, Development and Technological Innovation (IN- NPACTO subprogram), LTEExtreme project (IPT-2012-0525-430000), and the project TEC2014-59255-C3-3-R (ELISA).

REFERENCES

- [1] "Report - M2M Global Forecast & Analysis 2012-22," Machina Research, Tech. Rep., 2013.
- [2] J. Monserrat, J. Calabuig, A. Fernandez-Aguilella, and D. Gomez-Barquero, "Joint delivery of unicast and e-mbms services in lte networks," *Broadcasting, IEEE Transactions on*, vol. 58, no. 2, pp. 157–167, 2012.
- [3] "Introduction of the multimedia broadcast/multicast service (MBMS) in the radio access networks (RAN)-Stage 2 (Release 9)", 3GPP TS 25.346 v9.1.0, Mar 2010.
- [4] "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2", 3GPP TS 36.300 v11.7.0, Sep 2013.
- [5] R. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for ofdma-based systems: A survey," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 1, pp. 240–254, 2013.
- [6] G. Araniti, M. Condoluci, L. Militano, and A. Iera, "Adaptive Resource Allocation to Multicast Services in LTE Systems," *Broadcasting, IEEE Transactions on*, vol. 59, no. 4, pp. 658–664, Dec 2013.
- [7] A. d. I. Fuente, A. Garcia Armada, and R. P. Leal, "Joint Multicast/Unicast Scheduling with Dynamic Optimization for LTE Multicast Service," in *European Wireless 2014; 20th European Wireless Conference; Proceedings of*, May 2014, pp. 1–6.
- [8] G. Monghal, K. Pedersen, I. Kovacs, and P. Mogensen, "Qos oriented time and frequency domain packet schedulers for the utran long term evolution," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, 2008, pp. 2532–2536.
- [9] "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures", 3GPP TS 36.213 v11.4.0, Sep 2013.
- [10] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 2, pp. 678–700, 2013.

Asignación de Recursos Radio en Redes Industriales Inalámbricas Multisalto

Sergio Montero, Javier Gozalvez, Miguel Sepulcre

UWICORE, Ubiquitous Wireless Communications Research Laboratory, www.uwicare.umh.es

Universidad Miguel Hernández de Elche (UMH)

Avda. de la Universidad, s/n, 03202, Elche (España)

smontero@umh.es, j.gozalvez@umh.es, msepulcre@umh.es

Resumen- El uso de redes inalámbricas industriales supondrá una mejora de la productividad, la flexibilidad y la eficiencia de las fábricas del futuro debido a su alto potencial para facilitar el despliegue de nuevas aplicaciones. Algunas de las aplicaciones permitirán la incorporación de nodos móviles para su monitorización y control, tales como vehículos autónomos, robots móviles, e incluso trabajadores. Los estrictos requisitos de las aplicaciones industriales, unidos a la posibilidad de desplegar dispositivos móviles, requieren el diseño de algoritmos de asignación de recursos que coordinen las transmisiones de los diferentes nodos de la red para transportar de forma fiable y eficiente la información entre origen y destino. En este contexto, este artículo presenta y evalúa un nuevo algoritmo de asignación de recursos radio que emplea enlaces compartidos entre los nodos que forman una misma ruta para mejorar el rendimiento de las comunicaciones extremo a extremo. Los resultados obtenidos demuestran el potencial del algoritmo propuesto frente a algoritmos existentes, tanto en redes con nodos estáticos como en redes que incluyan nodos móviles.

Palabras Clave- Redes inalámbricas industriales, redes multisalto, asignación de recursos radio, movilidad.

I. INTRODUCCIÓN

En los últimos años, las redes inalámbricas industriales, (IWNs, *Industrial Wireless Networks*) han atraído la atención tanto de la industria como de la comunidad científica debido a las importantes ventajas que ofrecen frente a las tradicionales redes cableadas. Las tecnologías inalámbricas reducen el tiempo de instalación y mantenimiento, lo que mejora la flexibilidad y la capacidad de reconfiguración de la fábrica, permitiendo la conectividad con subsistemas móviles, dispositivos móviles o robots, entre otros [1]. Gracias a las IWNs, se podrán desarrollar nuevas aplicaciones, como la monitorización continua del estado de procesos y equipos críticos, o un seguimiento en tiempo real de los trabajadores u objetos. Sin embargo, el despliegue de las IWNs en las fábricas supone un importante desafío debido a la propia naturaleza del medio inalámbrico, puesto que está sujeto a condiciones de propagación variables y adversas, y a interferencias. Como resultado, las IWNs se han de diseñar para que sean capaces de garantizar en un entorno tan hostil los requisitos de fiabilidad y de latencia de las aplicaciones industriales. Esto puede ser especialmente crítico cuando los requisitos de las aplicaciones son elevados, y cuando los nodos son móviles.

Para hacer frente a estos retos, se han desarrollado los estándares industriales inalámbricos WirelessHART [2], ISA100.11a [3] e IEEE 802.15.4e [4]. Estos estándares comparten buena parte de sus características principales y algunos de los mecanismos empleados [5]. Por ejemplo,

todos se basan en un sistema centralizado de gestión de red que incrementa la capacidad de garantizar los niveles de fiabilidad y latencia requeridos por las aplicaciones de misión crítica soportadas [1]. Estos estándares se basan en la capa física y la capa MAC del estándar IEEE 802.15.4 operando en la banda de frecuencia ISM de 2,4 GHz. Además añaden un esquema de acceso múltiple por división en el tiempo (TDMA, *Time Division Multiple Access*), combinado con un esquema de salto en frecuencia (FH, *Frequency Hopping*) para mejorar la capacidad y la robustez de las comunicaciones inalámbricas. Por tanto, cada enlace de comunicación se identifica por un intervalo de tiempo y un canal de frecuencia. Todos los intervalos de tiempo tienen la misma duración, siendo lo suficientemente grandes como para acomodar la transmisión de un mensaje de datos y un mensaje de confirmación de la entrega desde el receptor.

El gestor de la red se encarga, entre otras tareas, de la asignación de los enlaces de comunicación a los nodos de la red para evitar que dos nodos transmitan en un mismo enlace, es decir, en un mismo intervalo de tiempo y canal de frecuencia. La asignación de los enlaces ha de permitir una transmisión fiable entre el origen y el destino, por lo que podrá necesitarse el empleo de nodos intermedios como retransmisores cuando los nodos origen y destino se encuentren fuera de sus rangos de cobertura. Además, para mejorar la fiabilidad de las comunicaciones es necesario permitir que cada nodo dentro de la ruta entre origen y destino pueda retransmitir los mensajes en caso de fallo, para lo cual es necesario asignar intervalos de tiempo adicionales en cada salto. En los algoritmos basados en TDMA, a cada nodo se ha de asignar un número limitado de enlaces para la transmisión de cada mensaje, y no siempre serán utilizados en función de cuántas retransmisiones sean necesarias en cada ocasión. Por tanto, el problema fundamental es cómo asignar de manera eficiente los enlaces, incluyendo los designados para las retransmisiones, para maximizar la probabilidad de que la información sea entregada de forma correcta y a tiempo a su destino. Este problema es especialmente relevante para la asignación de tráfico de datos periódico, ya que puede consumir una parte importante del ancho de banda. Este sería del tipo de tráfico generado por los sensores desplegados en las fábricas, ya que deben de reportar periódicamente sus medidas en un plazo de tiempo limitado y muy estricto. Para tratar de solventar este problema, este trabajo propone y evalúa un nuevo algoritmo de asignación de enlaces para redes inalámbricas industriales basado en un testigo virtual. El algoritmo propuesto emplea

enlaces compartidos entre los nodos que forman una misma ruta para mejorar el rendimiento de las comunicaciones extremo a extremo. Gracias a ello, es capaz de alcanzar compromiso entre la minimización del retardo y la maximización de la fiabilidad en la entrega del mensaje. Su rendimiento y eficiencia se ha evaluado en diferentes escenarios (incluyendo escenarios con nodos fijos y móviles) y comparado con algoritmos existentes.

El resto del artículo se organiza como se detalla a continuación. El segundo apartado presenta los algoritmos de asignación de enlaces más relevantes propuestos en la literatura, así como los empleados generalmente en comunicaciones inalámbricas industriales. El tercer apartado presenta el algoritmo propuesto. El apartado IV presenta las métricas de rendimiento y eficiencia utilizadas y el apartado V presenta los escenarios de evaluación y los resultados obtenidos. Finalmente el apartado VI presenta las conclusiones del estudio realizado.

II. ESTADO DEL ARTE

La asignación de enlaces en redes de sensores inalámbricas con transmisiones a través de múltiples saltos ha sido ampliamente estudiada en la literatura. Sin embargo, el hecho de emplear mayoritariamente algoritmos distribuidos, como los propuestos en [7], [8] o [9], impiden que éstos puedan ser aplicados a las IWNs, pues las redes industriales requieren una centralización de la toma de las decisiones. Generalmente, los algoritmos distribuidos propuestos se han centrado en minimizar el retardo medio de los mensajes mediante la asignación de enlaces libres de interferencia tan frecuentemente como sea posible. Para ello, se emplea la información que cada nodo obtiene de sus vecinos, de forma que varios nodos pueden auto-asignarse el mismo enlace para la comunicación. Sin embargo, los estándares inalámbricos industriales no permiten la reutilización espacial de un enlace (mismo intervalo de tiempo y mismo canal de frecuencia) para más de una transmisión, evitando que surjan interferencias. Los algoritmos centralizados, como [10] y [11], han sido enfocados en determinar la ruta completa entre origen y destino. En este tipo de algoritmos, un nodo centralizado es el encargado de realizar todas las asignaciones necesarias al conocer la información de todos los nodos de la red. Los enlaces están dispuestos de forma que se evitan totalmente las colisiones entre transmisiones, incrementando en gran medida el porcentaje de mensajes correctamente entregados, si bien suponen un alto nivel de señalización cuando se reorganiza la asignación de enlaces.

Recientemente se han realizado diferentes propuestas para los estándares inalámbricos industriales que tienen en cuenta el uso de diferentes canales de frecuencia [12][13]. Los algoritmos propuestos para los estándares industriales pueden ser aplicados a cualquier red en la que sea necesario emplear más de un salto para entregar el mensaje al destino; como por ejemplo una comunicación entre dos nodos que están fuera del rango de cobertura del otro o una comunicación en la que sea mejor utilizar nodos intermedios para entregar del mensaje al destino [14]. Además, en el caso de que tanto de la transmisión inicial como su retransmisión no se hayan completado de forma correcta, es posible establecer una ruta alternativa a través de otros nodos intermedios. También es necesario que una vez que la asignación de enlaces sea

establecida no haya colisiones o sobre-escucha, si bien sí es posible que haya ciertas escuchas innecesarias en ciertos enlaces, esto es, escuchas por parte de ciertos nodos en enlaces en los que no se transmite ningún mensaje destinado a ellos. Estas escuchas innecesarias, que han de ser evitadas pues consumen recursos, son debidas principalmente a que los nodos encargados de transmitir en dicho enlace no lo hacen. Esto puede ser debido a que ya haya transmitido el mensaje y mantenga activo un enlace redundante, o a que todavía no lo haya recibido.

El rendimiento y la eficiencia del algoritmo propuesto se han comparado con dos de los algoritmos de asignación de recursos en IWNs más empleados. En el primer algoritmo considerado, los enlaces dedicados a la transmisión de cada nodo (inicial y retransmisiones) se asignan antes que los enlaces dedicados a la transmisión del siguiente nodo en la ruta. Este algoritmo será referenciado en el resto del artículo con el nombre de Salto a Salto (SaS). Para una ruta compuesta por tres saltos (cuatro nodos) como la de la Fig. 1, la Fig. 2 muestra el ejemplo de asignación de enlaces al emplear el algoritmo SaS. La distribución de enlaces está además particularizada para el caso en que se asignan dos enlaces por cada salto del mensaje en la ruta, es decir, hay un enlace redundante en cada salto que puede ser utilizado si falla la primera transmisión. La principal ventaja de este algoritmo de asignación de enlaces es que el mensaje llega al destino si el número de errores de transmisión en cada salto es menor que el número de enlaces reservados a cada salto. La principal desventaja de este algoritmo es que los mensajes recibidos correctamente en el primer enlace de cada salto han de esperar a que transcurra el tiempo asignado a los enlaces dedicados a retransmisiones, con el consiguiente incremento del retardo extremo a extremo del mensaje. Además, los enlaces dedicados a retransmisiones sólo pueden ser utilizados por los nodos a los que previamente se les han asignados. Este algoritmo es utilizado por WirelessHART, así como por [15].

El segundo algoritmo de asignación de enlaces escogido para la comparativa se basa en la asignación de un enlace a cada uno de los saltos en la ruta desde el origen hasta el destino. Al término de la asignación anterior, se asigna un nuevo enlace a cada salto de la ruta para retransmitir el mensaje si es necesario, repitiendo el proceso de acuerdo al número de enlaces redundantes que vayan a ser asignados. Este algoritmo será referenciado en el resto del artículo con el nombre de Camino a Camino (CaC). La Fig. 2 muestra el ejemplo de asignación de enlaces al emplear el algoritmo CaC para la ruta presentada en la Fig. 1 cuando se asignan dos enlaces por cada salto. La principal ventaja de este algoritmo de asignación de enlaces es el bajo retardo obtenido en la entrega del mensaje cuando no hay errores en la transmisión. Sin embargo, cuando una de las transmisiones falla y se requiera una retransmisión, algunos de los enlaces dedicados a retransmisiones de los siguientes nodos en la ruta no pueden ser empleados en el caso de ser necesarios; por ejemplo, si falla la comunicación $A \rightarrow B$ en t_1 , los enlaces $B \rightarrow C$ en t_2 y $C \rightarrow D$ en t_3 no podrán utilizarse puesto que están asignados antes de que A pueda realizar la retransmisión en t_4 . De esta manera, aunque este algoritmo puede reducir el retardo del mensaje hasta el destino, supone una menor utilización de los enlaces dedicados a

retransmisiones, lo que se traduce en una reducción de la probabilidad de entregar correctamente el mensaje al destino. Este algoritmo es usado en [13] para su aplicación en redes que requieren fiabilidad y retardo predecibles, como sucede en las aplicaciones de automatización industriales.

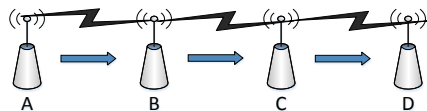


Fig. 1. Ruta compuesta por tres saltos que implica el uso de cuatro nodos para enviar un mensaje desde el nodo A al nodo D.

SaS	A → B	A → B	B → C	B → C	C → D	C → D
CaC	A → B	B → C	C → D	A → B	B → C	C → D
	t_1	t_2	t_3	t_4	t_5	t_6

Fig. 2. Ejemplo de asignación de enlaces empleando el algoritmo Salto a Salto (SaS) y el algoritmo Camino a Camino (CaC) para la transmisión de un mensaje en la ruta mostrada en la Fig. 1.

III. ALGORITMO DE ASIGNACIÓN DE RECURSOS PROPUESTO

En este trabajo, se propone un algoritmo de asignación de enlaces que aumenta la flexibilidad en el uso de los enlaces asignados a las retransmisiones, alcanzando un compromiso entre la minimización del retardo y la maximización de la fiabilidad en la entrega del mensaje. El algoritmo propuesto no sólo resulta beneficioso para despliegues de nodos estáticos como veremos en el apartado de evaluación, sino que también resulta muy beneficioso en redes industriales con nodos inalámbricos fijos y móviles. Esto es debido a que el algoritmo de asignación de enlaces propuesto permite un uso flexible de los enlaces dedicados a las retransmisiones por parte de los nodos que forman una misma ruta.

El algoritmo propuesto en este artículo, ARCO (Asignación de Recursos basado en la Compartición secuencial de enlaces), se fundamenta en el empleo de enlaces compartidos y un testigo virtual. Mediante ARCO, cada nodo ha de estar primero en estado de recepción para intentar recibir el mensaje, manteniéndose en dicho estado en los siguientes enlaces hasta la correcta recepción del mensaje. Una vez recibido el mensaje, el nodo tendrá que transmitirlo hacia el siguiente nodo en el siguiente enlace asignado para ello. La estructura básica de este algoritmo necesita un enlace por cada salto y un enlace por cada retransmisión permitida. El primer nodo de la ruta tiene asignado el primer enlace así como los siguientes R enlaces, donde R simboliza el número de retransmisiones permitidas. El segundo nodo tiene asignado el segundo enlace así como los siguientes R enlaces, repitiendo este algoritmo para cada uno de los nodos hasta el último nodo de la ruta que transmite el mensaje. Cada nodo, excepto el primero en la cadena de transmisión, necesita recibir el mensaje desde el anterior nodo en la ruta antes de usar alguno de sus enlaces. Una vez que un nodo transmite correctamente el mensaje al siguiente nodo en la ruta no necesita realizar más funciones.

La Fig. 3 muestra la distribución temporal de la técnica ARCO para la ruta presentada en la Fig. 1 ($S=3$ saltos) cuando se consideran un total de 6 enlaces asignados para el conjunto, al igual que los 6 enlaces considerados para SaS y CaC en la Fig. 2. A pesar de que el total de enlaces reservados es el mismo, ARCO permite que un nodo pueda

utilizar hasta tres enlaces para retransmisiones ($R=3$). Algunos de los enlaces se comparten por varios nodos, pero solo uno de ellos podrá transmitir simultáneamente, puesto que para ello necesitará haber recibido correctamente el mensaje del nodo que le precede en la ruta. La principal característica de ARCO es que permite la entrega del mensaje entre origen y destino siempre que la suma del número de transmisiones erróneas entre todos los saltos sea igual o inferior al número total de enlaces compartidos para retransmisiones (R). Es importante recordar que con el algoritmo SaS, la entrega del mensaje al destino sólo se produce si el número de transmisiones erróneas en cada salto es inferior al número de enlaces dedicados para retransmisión en dicho salto. En el algoritmo CaC, la entrega del mensaje al destino sólo se produce si el número total de transmisiones erróneas es inferior al número de enlaces asignados a cada nodo en la ruta. Así, con ARCO, tener R enlaces compartidos para retransmisiones permite a cada nodo disponer de hasta $R+1$ intentos para transmitir el mensaje hasta al siguiente nodo en la ruta. Si un nodo necesita $R+1$ intentos, los otros han de transmitir correctamente el mensaje en el primer intento para que el mensaje alcance el destino.

Otra de las ventajas de ARCO es que permite que cada nodo pueda aprovechar los enlaces compartidos para retransmisiones sólo cuando los necesite, a la vez que asegura que sólo se produce una transmisión en cada instante de tiempo. Así, ARCO se basa en que cada enlace compartido sólo puede ser empleado por el nodo que tenga pendiente transmitir el mensaje hasta el siguiente nodo en la ruta, actuando el propio mensaje como un testigo virtual que identifica quien puede utilizar el actual enlace para la transmisión del mensaje.

A → B	A → B	A → B	A → B		
	B → C	B → C	B → C	B → C	
		C → D	C → D	C → D	C → D
t_1	t_2	t_3	t_4	t_5	t_6

Fig. 3. Ejemplo de asignación de enlaces empleando el algoritmo ARCO para la transmisión de un mensaje a través de la ruta de la Fig. 1 particularizada para el caso de 3 enlaces compartidos para retransmisiones.

Para entender mejor su funcionamiento la Fig. 4 muestra un ejemplo empleando ARCO que se corresponde con el caso en el que hay una transmisión errónea en el primer salto y otra en el tercer salto (marcadas con una cruz). El resto de transmisiones son correctas y se han señalado con un círculo. Así, en este ejemplo la transmisión entre A y D alcanza con éxito el destino tras cinco enlaces (uno para cada uno de los 3 saltos y dos para las retransmisiones realizadas).

A × B	A → B	A → B	A → B		
	B → C	B → C	B → C	B → C	
		C → D	C × D	C → D	C → D
t_1	t_2	t_3	t_4	t_5	t_6

Fig. 4. Ejemplo de funcionamiento de ARCO cuando se produce una transmisión errónea en el primer y tercer salto al considerar $S=3$ y $R=3$.

IV. MÉTRICAS DE RENDIMIENTO

Las métricas de rendimiento utilizadas en este trabajo permiten evaluar la probabilidad de que el mensaje sea correctamente entregado al destino, el retardo con el que se

entrega y la eficiencia de los algoritmos en términos de enlaces utilizados, enlaces bloqueados, intentos de recepción innecesarios y consumo energético.

A. Fiabilidad

La fiabilidad de los diferentes algoritmos se ha evaluado mediante la métrica PDR_{E2E} (*End-to-end Packet Delivery Ratio*). Dicha métrica se calcula como el ratio de mensajes correctamente entregados al destino respecto al número total de mensajes generados. Considerando una PDR estacionaria para cada salto (PDR_i es la ratio de entrega de mensajes en el salto i), la PDR_{E2E} puede ser expresada mediante una expresión analítica para los diferentes algoritmos. En este sentido, la PDR_{E2E} puede ser estimada para SaS como:

$$PDR_{E2E} = \prod_{i=1}^S (1 - (1 - PDR_i)^{E_i}) \quad (1)$$

donde S es la cantidad de saltos del mensaje hasta el destino y E_i es el número de enlaces dedicados al nodo encargado de la transmisión en el salto i usando el algoritmo SaS.

La PDR_{E2E} también puede ser expresada para el algoritmo CaC, tal y como se muestra en la Ec. 2. En esta ecuación, la variable E representa el número de veces que un enlace es dedicado a la transmisión en cada salto. Así, $E=2$ indica que el proceso de asignar un enlace a cada uno de los saltos se realiza dos veces, mientras $E=3$ repite el proceso 3 veces. Esta ecuación podría extenderse fácilmente para valores de E mayores.

$$PDR_{E2E} = \begin{cases} \left(\prod_{i=1}^S PDR_i \right) \cdot \left(1 + \left(\sum_{i=1}^S (1 - PDR_i) \right) \right) & \text{si } E=2 \\ \left(\prod_{i=1}^S PDR_i \right) \cdot \left(1 + \left(\sum_{i=1}^S (1 - PDR_i) \right) + \left(\sum_{i=1}^S \sum_{j=1}^S (1 - PDR_i)(1 - PDR_j) \right) \right) & \text{si } E=3 \end{cases} \quad (2)$$

Finalmente, una expresión de la PDR_{E2E} puede obtenerse para ARCO. La correspondiente expresión, presentada en la Ec. 3, depende del número total de enlaces de retransmisión compartidos (R), así como del número de saltos necesarios del mensaje para alcanzar el destino (S).

$$PDR_{E2E} = \begin{cases} \left(\prod_{i=1}^S PDR_i \right) \cdot \left(1 + \left(\sum_{i=1}^S PDR_i \right) \right) & \text{si } R=1 \\ \left(\prod_{i=1}^S PDR_i \right) \cdot \left(1 + \left(\sum_{i=1}^S PDR_i \right) + \left(\sum_{i=1}^S \sum_{j=1}^S PDR_i \cdot PDR_j \right) \right) & \text{si } R=2 \\ \left(\prod_{i=1}^H PDR_i \right) \cdot \left(1 + \left(\sum_{i=1}^S PDR_i \right) + \left(\sum_{i=1}^S \sum_{j=1}^S PDR_i \cdot PDR_j \right) \right) & \text{si } R=3 \end{cases} \quad (3)$$

B. Retardo extremo a extremo

En IWNs, el retardo con el que se entrega cada mensaje es de especial importancia para el correcto funcionamiento de las aplicaciones. En este contexto, el retardo medio extremo a extremo D_{E2E} (*End-to-end Delay*) es definido como el tiempo medio transcurrido desde que el mensaje es enviado por el nodo origen hasta que es entregado al nodo destino. En este estudio se considera que todos los enlaces dedicados a la transmisión del mensaje son asignados de manera consecutiva siguiendo el algoritmo correspondiente. Es importante resaltar que solo los mensajes que son correctamente entregados al destino son tenidos en cuenta para obtener el retardo medio. Esta consideración se ha realizado ya que generalmente los mensajes periódicos no entregados en los recursos asignados, no son retransmitidos

en los recursos asignados al siguiente periodo, puesto que en este último periodo se enviará un nuevo mensaje con información del sensor actualizada.

C. Enlaces utilizados

La primera métrica definida para analizar la eficiencia de los diferentes algoritmos es el uso medio de todos los enlaces dedicados para la transmisión de cada mensaje hasta su destino. En función del algoritmo de asignación de enlaces y los errores que se produzcan en las transmisiones, muchos de los enlaces asignados pueden quedarse sin utilizar. Una buena asignación de enlaces explotará al máximo los enlaces asignados, obteniendo una alta utilización de los mismos. Esta métrica no tiene en cuenta si el mensaje ha sido o no correctamente entregado al destino ya que es una medida del uso general de los recursos dedicados a cada algoritmo. Así, un mayor porcentaje de enlaces utilizados supone que una mayor cantidad de entre los enlaces dedicados, son utilizados para intentar que el mensaje alcance el destino.

D. Enlaces bloqueados

Otra métrica definida para analizar la eficiencia es el número de enlaces bloqueados, esto es, el número de enlaces en los que un nodo ha de estar en transmisión o recepción por cada mensaje que se quiera transmitir entre origen y destino. Es importante destacar que los nodos asignados a un enlace no pueden ser asignados a la transmisión o recepción de otros mensajes en otros enlaces en el mismo instante de tiempo y en otro canal de frecuencia ya que no se permiten duplicidades de funciones radio en un mismo instante de tiempo empleando dos canales de frecuencia distintos.

E. Intentos de recepción innecesarios

La tercera métrica de rendimiento que analiza la eficiencia son los Intentos de Recepción Innecesarios o *IRI* de cada mensaje entregado al destino. Esta métrica evalúa la cantidad de veces que un nodo permanece en modo recepción en un enlace en el que no hay ningún mensaje destinado a él. Esto puede ser así porque dicho nodo no es el destino del mensaje, o bien porque no hay ningún nodo transmitiendo un mensaje en ese enlace. Esta métrica representa una medida de la ineficiencia de los algoritmos de asignación de enlaces en términos de energía consumida ineficientemente, pues el nodo podría haber permanecido en un estado de reposo en dicho enlace o podría haber sido asignado a otras funciones.

F. Consumo energético

Este estudio también ha considerado la evaluación del consumo medio de energía por cada mensaje entregado al destino, E_{PME} . Esta métrica es más justa que el consumo total de energía, puesto que tiene en cuenta que el número de transmisiones realizadas puede ser variable en función de los ratios de entrega de cada algoritmo de asignación. Por ejemplo, si simplemente midiéramos la energía total consumida, un algoritmo con un ratio de entrega reducido podría parecer tener un consumo muy bajo, puesto que los mensajes no llegan a ser transmitidos hasta el destino.

V. EVALUACIÓN DEL RENDIMIENTO

El rendimiento de los diferentes algoritmos de asignación de enlaces ha sido evaluado en tres escenarios diferentes, cada uno de ellos con características que los hacen

particularmente relevantes. Todos los escenarios consideran que los mensajes son enviados a su destino a través de dos nodos intermedios (tres saltos) y se representan en la Fig. 5. El escenario 1 considera tres saltos con la misma PDR , es decir, $PDR_{AB}=PDR_{BC}=PDR_{CD}$. Este escenario es empleado para analizar el rendimiento de los diferentes algoritmos de asignación de enlaces mediante un barrido de todos los posibles valores PDR [0-1], lo que permite obtener unas primeras estimaciones del rendimiento de los algoritmos.

Para simular unas condiciones de funcionamiento realistas, en los escenarios 2 y 3 se simulan nodos que implementan el transceptor XDM2510H [16], transceptor radio compatible con los estándares industriales como WirelessHART [2] en la banda de frecuencia de los 2.4 GHz. El transceptor, alimentado con 3.6V, tiene un consumo de 18mA y 6mA cuando está en modo de transmisión y recepción, respectivamente. El transmisor tiene un consumo de 8.5μA cuando la radio está desactivada, consumo de energía que es despreciado. La potencia de transmisión es de 8dBm con una sensibilidad de recepción de -90dBm. Para evaluar el consumo de energía se considera que los nodos transmiten mensajes de datos de 133 Bytes y cada mensaje es confirmado en cada salto con un mensaje de confirmación de 26 Bytes. Además, con el objetivo de simular unas condiciones de propagación realistas en los escenarios 2 y 3 se ha implementado en MATLAB el modelo de propagación radio propuesto en [17] para entornos industriales. Algunos de los principales parámetros de dicho modelo son: $d_0=15m$ (distancia de referencia), $PL(d_0)=71.84dBm$ (perdidas a la distancia de referencia), $n=2.16$ (exponente de pérdidas), $\sigma=8.13dBm$ (desviación estándar del desvanecimiento) y $G_t=G_r=0dB$ (ganancia de la antena de transmisión y recepción). Este estudio ha considerado que una transmisión es recibida correctamente por el destino si el nivel de RSSI en el receptor es superior a -90dBm, ya que no hay posibilidad de dos transmisiones en un mismo enlace.

El escenario 2, mostrado en la Fig. 5, se corresponde con un escenario en el que los cuatro nodos permanecen estáticos y separados una cierta distancia. Dos separaciones entre nodos d_{AB} , d_{BC} y d_{CD} , han sido analizadas en este escenario, 50m y 150m, cubriendo todas las posibles combinaciones (8 casos) mientras se mantiene fijo el número de saltos. El escenario 3, en cambio, representa un escenario en el que el nodo A se mueve dentro del rango de cobertura de B. El nodo A siempre utiliza a B para transmitir los mensajes hasta D mientras se desplaza. En este escenario, la máxima separación permitida entre A y B (d_{AB}) es de 150 metros. Sin embargo, y de forma similar al escenario 2, se han analizado dos diferentes distancias entre nodos d_{BC} y d_{CD} , (50m y 150m), cubriendo las 4 posibles combinaciones mientras se mantiene fijo el número de saltos.

Los diferentes algoritmos evaluados han empleado la asignación de enlaces presentada en la Fig. 2 y la Fig. 3. Todos los algoritmos evaluados han dedicado 6 enlaces para cada mensaje que se quiera transmitir entre origen y destino. Además, con el objetivo de ver el efecto de las retransmisiones, también se ha evaluado un algoritmo que no emplea retransmisiones de los mensajes. Este algoritmo, identificado como Nrtx, sólo necesita tres enlaces por cada mensaje que se quiera transmitir entre origen y destino. Se ha creído conveniente su inclusión en la comparativa realizada

al representar el algoritmo más básico de transmisión multi-salto. El rendimiento de los diferentes algoritmos de asignación de enlaces ha sido evaluado mediante extensas simulaciones siguiendo la configuración y los escenarios descritos en este apartado. A continuación se presentan los resultados de la evaluación realizada para cada uno de los tres escenarios considerados.

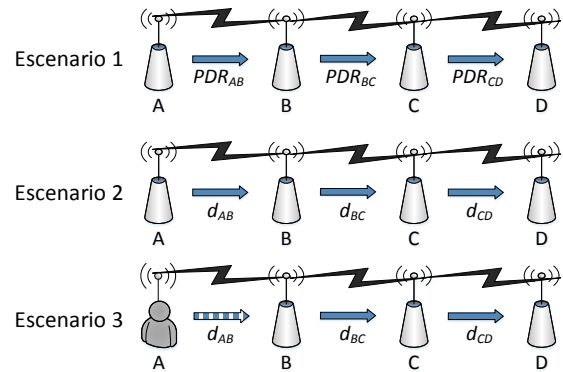


Fig. 5. Escenarios considerados para la evaluación de los diferentes algoritmos de asignación de enlaces.

A. Resultados - escenario 1

La Fig. 6a representa la probabilidad de que un mensaje sea correctamente entregado al nodo de destino (PDR_{E2E}) en el escenario 1, en el que la PDR de los 3 enlaces es la misma e igual a PDR_i . Los resultados obtenidos muestran que la PDR_{E2E} del algoritmo propuesto supera siempre a la de los otros algoritmos. El rendimiento del algoritmo propuesto es especialmente superior para valores de PDR_i intermedios [40-70%]. El rendimiento obtenido por el algoritmo de SaS es ligeramente superior al obtenido por el algoritmo CaC ya que para el escenario evaluado, el algoritmo SaS permite un error por cada salto mientras que el algoritmo CaC sólo permite un error en uno de los saltos. Con la evaluación de esta métrica se confirma que el empleo de enlaces compartidos con el objetivo de usar los recursos inalámbricos de una forma más eficiente logra una mayor probabilidad de entregar correctamente el mensaje al destino. En la Fig. 6b se muestra el retardo medio entre origen y destino de los mensajes (D_{E2E}) en función de la PDR_i , cuando los tres saltos tienen la misma PDR . En esta figura, el retardo se ha representado en número de enlaces. Es importante recordar, que en el cómputo de D_{E2E} sólo se tiene en cuenta los mensajes que se entregan correctamente al destino. Los resultados obtenidos muestran que el algoritmo ARCO propuesto obtiene el menor D_{E2E} de entre las diferentes técnicas que disponen de enlaces para retransmisiones. El D_{E2E} obtenido por el algoritmo sin retransmisiones (Nrtx) es igual a 3 enlaces ya que en este algoritmo todos los mensajes que son entregados al destino lo hacen en el tercer enlace. Sin embargo hay que tener en cuenta la baja PDR_{E2E} obtenida en la Fig. 6a por esta técnica, lo que se traduce en una baja cantidad de mensajes entregados con un retardo de tres enlaces para valores medios y bajos de PDR_i . Para valores altos de PDR_i , se puede apreciar que el D_{E2E} obtenido por los algoritmos ARCO y CaC se aproximan al retardo obtenido por la técnica que no emplea retransmisiones. En cambio, el algoritmo SaS supone un retardo mínimo de al menos cinco enlaces. Con la evaluación del retardo se confirma que el uso

de enlaces compartidos mediante testigo virtual logra reducir el retardo de la transmisión del mensaje hasta el destino.

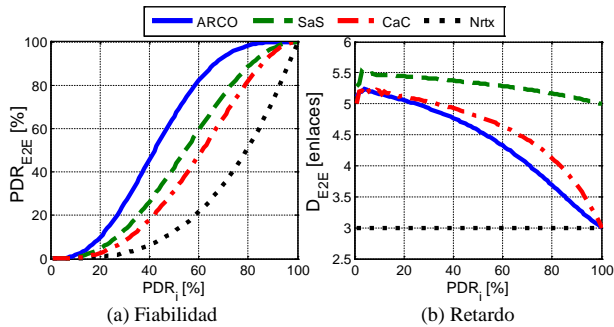


Fig. 6. Fiabilidad y Retardo en función de la PDR_i de todos los saltos para los diferentes algoritmos de asignación de enlaces cuando se evalúa el escenario 1 con seis enlaces dedicados y tres saltos con la misma PDR_i .

La eficiencia es también un aspecto relevante a analizar. La Fig. 7a muestra la utilización media de todos los enlaces al emplear cada algoritmo. Los resultados presentados muestran que ARCO realiza una mejor utilización de los enlaces asignados que SaS y CaC independientemente del valor de PDR_i considerado. Además, para valores bajos y medios de PDR_i , que se corresponden con situaciones en las que muy pocos mensajes alcanzan el segundo salto de la ruta, el rendimiento del algoritmo propuesto es muy alto. Esto indica que ARCO permite una alta utilización de los enlaces en los primeros saltos, con el consiguiente incremento de mensajes que alcanzan el destino. Sin embargo, para valores altos de PDR_i , la mayor utilización de los enlaces es obtenida por Nrtx. Esto es algo lógico, puesto que la técnica sin retransmisiones sólo dedica 3 enlaces y en valores altos de PDR_i el número medio de retransmisiones necesarias comienza a ser muy reducido. La Fig. 7b muestra que a pesar de los notables beneficios de ARCO, se produce un mayor número de Intentos de Recepción Innecesarios (IRI). Esto es así porque cada retransmisión del mensaje supone un retardo en la recepción del mensaje que actúa como testigo, incrementando el número de enlaces que un nodo está en recepción en un enlace en el que no tiene destinado un mensaje.

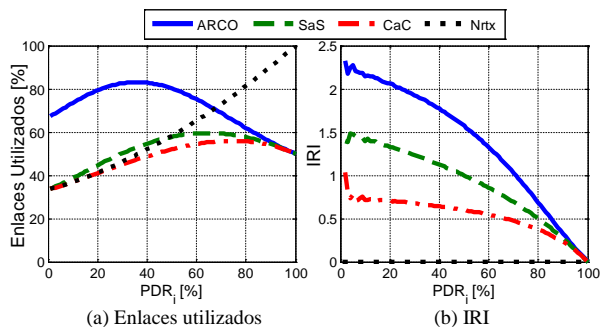


Fig. 7. Eficiencia en función de la PDR_i de todos los saltos para los diferentes algoritmos cuando se evalúa el escenario 1 con seis enlaces dedicados y tres saltos con la misma PDR_i .

La Tabla I presenta el número de enlaces que un nodo permanece bloqueado para estar en transmisión o recepción por cada mensaje que se quiera transmitir entre origen y destino. El algoritmo ARCO es el que presenta el mayor número de enlaces bloqueados. En los nodos origen y destino prácticamente se duplica el número de enlaces bloqueados con ARCO con respecto a las técnicas SaS y CaC. Sin

embargo, en los dispositivos de los saltos intermedios (B y C) sólo se incrementa en una unidad el número de enlaces bloqueados con ARCO respecto a SaS y CaC.

TABLA I
ENLACES BLOQUEADOS PARA TODOS LOS ESCENARIOS EVALUADOS

Nodo	ARCO	SaS	CaC	Nrtx
A	4	2	2	1
B	5	4	4	2
C	5	4	4	2
D	4	2	2	1

La Fig. 8 muestra la métrica E_{PMD} (energía por cada mensaje entregado al destino). Como puede observarse, los niveles obtenidos son muy similares para los diferentes algoritmos analizados, aunque es ligeramente inferior el obtenido con ARCO respecto a los otros algoritmos. Este mejor rendimiento por parte de ARCO es consecuencia de que la mayor probabilidad de que un mensaje sea correctamente entregado al nodo de destino compensa el incremento de energía ocasionado por el mayor número de intentos de recepción innecesarios de ARCO (Fig. 7b).

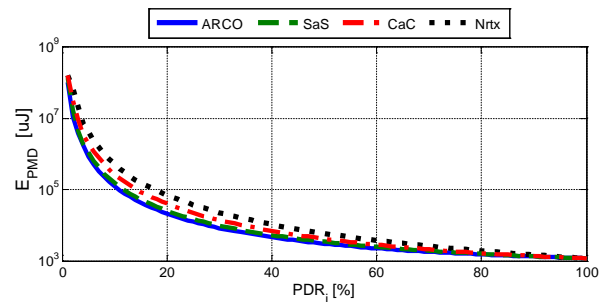


Fig. 8. Consumo de energía en función de la PDR_i de todos los saltos para los diferentes algoritmos cuando se evalúa el escenario 1 con seis enlaces dedicados y tres saltos con la misma PDR_i .

B. Resultados - escenario 2

Considerando condiciones de funcionamiento y de propagación realistas, en el escenario 2 se han evaluado 8 casos diferentes que corresponden a las 8 diferentes situaciones en las que los nodos de la ruta están separados 50 o 150 metros. Los diferentes casos se representan mediante $[d_{AB}, d_{BC}, d_{CD}]$, de forma que la primera distancia se corresponde a la separación entre A y B, la segunda distancia a la separación entre B y C y la tercera se corresponde a la separación entre C y D. Es importante resaltar que los resultados obtenidos muestran las mismas tendencias generales que las obtenidas en el escenario 1. Sin embargo, debido a que cada caso tiene un número distinto de saltos de buena o mala calidad, se va a hacer especial insistencia en las implicaciones que esto tiene en los diferentes resultados obtenidos. Los saltos de baja calidad se corresponden a saltos entre dispositivos separados 150 metros, mientras que los saltos de alta calidad se corresponden a saltos entre dispositivos separados 50 metros.

En primer lugar, la Tabla II presenta la PDR_{EZE} para los diferentes casos y algoritmos considerados. En este escenario, los resultados obtenidos muestran que en caso de que haya al menos un salto de 150 metros, el protocolo propuesto obtiene una PDR_{EZE} notablemente mayor que la de los algoritmos SaS y CaC. Además, los resultados obtenidos muestran que el valor de la PDR_{EZE} alcanzado por los

diferentes algoritmos depende de la cantidad de saltos de 150 metros, pero no así de la posición que éstos ocupan en la ruta.

TABLA II
PDR_{EZE} DEL ESCENARIO 2

$[d_{AB}, d_{BC}, d_{CD}]$	ARCO	SaS	CaC	Nrtx
[50, 50, 50]	99,9%	99,7%	99,3%	90,2%
[50, 50, 150]	99,2%	91,5%	90,2%	66,5%
[50, 150, 50]	99,2%	91,6%	90,1%	66,5%
[150, 50, 50]	99,2%	91,5%	90,1%	66,5%
[150, 150, 50]	97,2%	84,1%	78,9%	49,1%
[150, 50, 150]	97,2%	84,0%	78,9%	49,1%
[50, 150, 150]	97,2%	84,0%	78,9%	49,1%
[150, 150, 150]	93,9%	77,2%	67,4%	36,2%

La Tabla III presenta el retardo D_{EZE} para los diferentes casos y algoritmos considerados. En este escenario, los resultados obtenidos vuelven a mostrar que ARCO obtiene un retardo notablemente inferior a los otros algoritmos con retransmisiones. Los resultados obtenidos muestran que sólo para el algoritmo SaS, el D_{EZE} no depende de la cantidad de saltos de 150 metros y sí de su posición en la ruta, pues sólo varía si el último salto es de 50 o 150 metros. Así, en el algoritmo SaS, el retardo medio es de 5,03 enlaces y 5,22 para los casos con un último salto de 50 o 150 metros, respectivamente. En cambio, para el algoritmo ARCO, el retardo medio es menor a 4 enlaces y depende del número de saltos de 150 metros, siendo menor a menor cantidad de saltos de 150 metros haya.

TABLA III
 D_{EZE} [ENLACES] DEL ESCENARIO 2

$[d_{AB}, d_{BC}, d_{CD}]$	ARCO	SaS	CaC	Nrtx
[50, 50, 50]	3,10	5,03	3,28	3,00
[50, 50, 150]	3,44	5,22	3,79	3,00
[50, 150, 50]	3,44	5,03	3,79	3,00
[150, 50, 50]	3,44	5,03	3,79	3,00
[150, 150, 50]	3,74	5,03	4,14	3,00
[150, 50, 150]	3,74	5,22	4,13	3,00
[50, 150, 150]	3,74	5,22	4,14	3,00
[150, 150, 150]	3,99	5,22	4,39	3,00

La Tabla IV presenta la utilización media de todos los enlaces dedicados. Los resultados obtenidos muestran que ARCO vuelve a ser el que mayor utilización de los enlaces asignados realiza, de entre aquellos con retransmisores. Además, el incremento de enlaces utilizados es mayor en aquellos casos con mayor número de saltos de 150 metros, estando para el último caso muy cerca del rendimiento obtenido por Nrtx. La cantidad de enlaces bloqueados no cambia con respecto al escenario 1, por lo que vuelven a ser válidos los resultados de la Tabla I, puesto que no dependen de las condiciones de propagación.

TABLA IV
PORCENTAJE DE ENLACES UTILIZADOS DEL ESCENARIO 2

$[d_{AB}, d_{BC}, d_{CD}]$	ARCO	SaS	CaC	Nrtx
[50, 50, 50]	51,7%	51,6%	51,6%	96,7%
[50, 50, 150]	57,7%	55,9%	55,5%	96,7%
[50, 150, 50]	57,6%	54,5%	54,1%	88,5%
[150, 50, 50]	57,5%	53,1%	52,7%	80,0%
[150, 150, 50]	62,8%	55,7%	53,8%	74,0%
[150, 50, 150]	63,1%	56,9%	55,6%	80,1%
[50, 150, 150]	63,2%	58,3%	57,0%	88,5%
[150, 150, 150]	68,0%	59,2%	55,9%	74,0%

Por último, la Tabla V y la Tabla VI presentan el número de Intentos de Recepción Innecesarios (IRI) y la energía por cada mensaje entregado al destino (E_{PME}), respectivamente. En este escenario, en general el IRI es mayor para los casos en los que los saltos de 150 metros (aquellos con peor calidad) se sitúan en los primeros saltos de la ruta. En el caso particular del algoritmo propuesto es debido al hecho de que cada transmisión errónea implica un intento de recepción desperdiciado por cada salto que reste por realizar en la ruta. Sin embargo, de manera similar al escenario 1, cuando se compara la energía por cada mensaje entregado al destino (Tabla VI), los resultados obtenidos muestran un consumo energético muy similar entre los diferentes algoritmos. Sin embargo, y como era de esperar, los casos con mayor número de saltos de 150 metros son aquellos que obtienen un mayor consumo de energía por cada mensaje entregado al destino. Además, la reducción del consumo de energía conseguida por el mecanismo propuesto es mayor para aquellos casos con mayor número de baja calidad (saltos de 150 metros).

TABLA V
INTENTOS DE RECEPCIÓN INNECESARIOS DEL ESCENARIO 2

$[d_{AB}, d_{BC}, d_{CD}]$	ARCO	SaS	CaC	Nrtx
[50, 50, 50]	0,10	0,10	0,09	0
[50, 50, 150]	0,10	0,10	0,07	0
[50, 150, 50]	0,44	0,29	0,26	0
[150, 50, 50]	0,78	0,48	0,45	0
[150, 150, 50]	1,05	0,67	0,54	0
[150, 50, 150]	0,73	0,48	0,38	0
[50, 150, 150]	0,42	0,29	0,22	0
[150, 150, 150]	0,99	0,67	0,46	0

TABLA VI
ENERGÍA POR CADA MENSAJE ENTREGADO [μ J] DEL ESCENARIO 2

$[d_{AB}, d_{BC}, d_{CD}]$	ARCO	SaS	CaC	Nrtx
[50, 50, 50]	1225	1226	1229	1268
[50, 50, 150]	1377	1444	1457	1721
[50, 150, 50]	1388	1422	1434	1590
[150, 50, 50]	1400	1401	1413	1455
[150, 150, 50]	1572	1613	1661	1843
[150, 50, 150]	1566	1634	1700	1972
[50, 150, 150]	1554	1658	1725	2154
[150, 150, 150]	1758	1865	2021	2497

C. Resultados - escenario 3

El escenario 3 considera condiciones de funcionamiento y de propagación realistas y además incluye un nodo móvil. En concreto, el escenario 3 representa la situación en la que el nodo A se mueve dentro del rango de cobertura de B (separación de hasta 150 metros) para 4 casos en los que la distancia entre B y C y entre C y D puede ser de 50 o 150 metros. El movimiento del nodo móvil se realiza siguiendo el modelo *random waypoint*, mediante el cual el nodo de forma periódica selecciona aleatoriamente una posición y se dirige hacia ella a 1 m/s. Los diferentes casos se representan mediante $[d_{BC}, d_{CD}]$, de forma que la primera distancia se corresponde a la separación entre B y C y la segunda distancia se corresponde a la separación entre C y D. Los resultados obtenidos y presentados en la Fig. 9 muestran tendencias similares a los resultados obtenidos para los dos escenarios anteriores. Las barras en las diferentes figuras representan los valores promedio, mientras que las líneas verticales representan el intervalo entre los percentiles 5 y 95% de las diferentes métricas evaluadas. Sin embargo, los

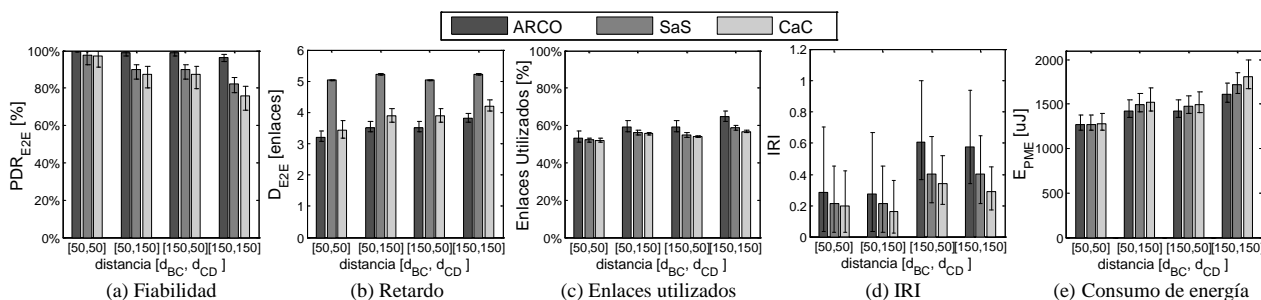


Fig. 9. Resultados obtenidos para los diferentes casos del escenario 3 a evaluar los distintos algoritmos. Las barras de la figura representan el valor medio mientras que las líneas verticales representan el intervalo entre el percentil 5% y 95%.

resultados de este escenario confirman la significativa mejora de rendimiento con el algoritmo propuesto al evaluar un escenario en el que el mensaje es originado por un dispositivo móvil. Así, en este escenario 3, el algoritmo propuesto obtiene la mayor PDR_{EZE} (Fig. 9a), el menor retardo (Fig. 9b), el mayor porcentaje de enlaces utilizados (Fig. 9c) y el menor consumo de energía por cada mensaje entregado al destino (Fig. 9e) incluso para los casos en los que los saltos entre los dispositivos estáticos son de sólo 50 metros. Además, saltos con mayor separación entre los nodos estáticos implica mayor incremento de rendimiento en el algoritmo ARCO, excepto de nuevo por el IRI (Fig. 9d).

VI. CONCLUSIONES

Este artículo ha propuesto y evaluado un nuevo algoritmo de asignación de enlaces para redes inalámbricas industriales con múltiples saltos. El algoritmo propuesto emplea enlaces compartidos entre nodos de una misma ruta con el objetivo de usar los recursos radio de una forma más eficiente gracias a una selección inteligente de los nodos que deben transmitir en cada instante de tiempo. El estudio presentado incluye el modelo analítico de la PDR_{EZE} de los diferentes algoritmos evaluados, modelo que puede ser empleado para obtener de manera sencilla una estimación de la PDR_{EZE} obtenida al emplear cada algoritmo. Los diferentes algoritmos de asignación de enlaces han sido evaluados mediante simulación teniendo en cuenta condiciones realistas de funcionamiento, incluyendo despliegues con nodos móviles y estáticos. Los resultados obtenidos han mostrado que el algoritmo propuesto mejora significativamente la fiabilidad y el retardo de las comunicaciones extremo a extremo. El incremento de rendimiento se logra a costa de un emplear un mayor número de enlaces bloqueados por cada nodo y de intentos de recepción innecesarios. Sin embargo, el algoritmo propuesto consume una menor energía por cada mensaje entregado al destino. Los resultados obtenidos demuestran también que el algoritmo propuesto se adapta especialmente a los despliegues con dispositivos móviles. Esto es debido a la selección dinámica de los nodos que deben transmitir en cada enlace, lo que permite aprovechar los enlaces extra asignados para realizar las retransmisiones en aquellos saltos en que sean necesarios.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad y los fondos FEDER (TEC2014-57146-R), y la Generalitat Valenciana (ACIF/2013/060).

REFERENCIAS

- [1] A. Kim, et al., "When HART goes wireless: Understanding and implementing the WirelessHART standard," *Proc. 13th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2008)*, Germany, Sep. 2008.
- [2] IEC 62591 Ed. 1.0: Industrial communication networks -Wireless communication network and communication profiles-WirelessHART™, IEC, 2010.
- [3] Wireless Systems for Industrial Automation: Process Control and Related Applications, ISA-100.11a-2009 Standard, 2009.
- [4] IEEE 802.15 WPAN™ Task Group 4e (TG4e). In <http://www.ieee802.org/15/pub/TG4e.html>.
- [5] S. Petersen et al., "WirelessHART Versus ISA100.11a: The Format War Hits the Factory Floor," *IEEE Industrial Electronics Magazine*, vol. 5, no. 4, pp. 23-34, Dec. 2011.
- [6] S. Nelson et al., "Spatial TDMA: A collision-free multihop channel access protocol," *IEEE Trans. Commun.*, vol. 33, pp. 934-944, 1985.
- [7] P. Zand, et al., "A Distributed Scheduling Algorithm for Real-time (D-SAR) Industrial Wireless Sensor and Actuator Networks," *Proc. 17th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA'12)*, Poland, Sept. 2012.
- [8] A. Saifullah, et al., "Distributed Channel Allocation Protocols for Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol.25, no.9, pp.2264,2274, Sept. 2014.
- [9] X. Zhang, et al., "CC-TDMA: Coloring- and Coding-based multi-channel TDMA scheduling for wireless ad hoc networks," *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC 2007)*, China, Mar. 2007, pp. 133-137.
- [10] Y. Li, et al., "Optimal link scheduling for delay-constrained periodic traffic over unreliable wireless links," *Proc. 33th IEEE Conf. Comput. Commun. (INFOCOM'14)*, Canada, April 2014, pp. 1465-1473.
- [11] S. Zhang, et al., "A highly reliable link scheduling strategy for WirelessHART networks," *Proc. 2013 IEEE International Conf on Advanced Techn. for Commun. (ATC)*, Vietnam, Oct 2013, pp. 39-43.
- [12] P. Suriyachai, et al., "Implementation of a MAC protocol for QoS support in wireless sensor networks," *Proc. IEEE International Conference on Pervasive Computing and Communications (PerCom)*, USA, March 2009.
- [13] S. Girs, et al., "The effects of relay behavior and position in wireless industrial networks", *Proc. IEEE Int. Workshop Factory Commun. Syst.*, France, May 2012, pp. 183-190.
- [14] K. Dang, et al., "A Graph Route-Based Superframe Scheduling Scheme in WirelessHART Mesh Networks for High Robustness," *Wireless Personal Communications*, vol. 71, pp. 2431-2444, 2013.
- [15] XDM2510H, 2.4 GHz IEEE 802.15.4 WirelessHART compliant radio module, RF Monolithics, Mar. 2011.
- [16] E. Tanghe, et al., "The industrial indoor channel: large-scale and temporal fading at 900, 2400, and 5200 MHz," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2740-2751, Jul. 2008.

Effect of Network Coding and Multi-hop Beaconing on the Channel Load of Vehicular Networks

Miguel Sepulcre, Javier Gozalvez, M. Carmen Lucas-Estañ, Baldomero Coll-Perales,
Sergio Montero, Jose Ramon Gisbert, Alejandro Moraleda

UWICORE, Ubiquitous Wireless Communications Research Laboratory, www.uwicore.umh.es
Universidad Miguel Hernández de Elche (UMH). Avda. de la Universidad, s/n, 03202, Elche (España)
{msepulcre, j.gozalvez, m.lucas, bcoll, smontero, jgisbert, amoraleda}@umh.es

Abstract- Vehicular networks are based on the periodic exchange of beacons on the so called control channel using the IEEE 802.11p technology. The critical nature of this channel has fostered significant efforts in the research and standardization communities to design congestion control protocols that dynamically adapt the transmission parameters to control the channel load. In this context, network coding has been demonstrated to improve the bandwidth efficiency in many different types of networks, but limited efforts have been conducted to date for its application to vehicular networks. This paper presents our first results on the use of network coding for the beaconing process in vehicular networks. The results obtained show the potential of network coding to reduce the congestion control problem and motivate the design of efficient multi-hop beaconing algorithms that exploit network coding. Network coding could present a significant impact on the design of the Decentralized Congestion Control (DCC) module that is currently being studied under the ETSI Technical Committee on ITS.

Keywords- Network coding, multi-hop beaconing, congestion control, vehicular networks, connected vehicle

I. INTRODUCTION

Network coding can be considered a routing technique with which intermediate network nodes process (encode) the information to be forwarded. In particular, intermediate nodes can use network coding to combine multiple packets into a single coded packet, and transmit the coded packet instead of transmitting each packet separately. As a result, network coding can improve resource efficiency (e.g. bandwidth and power) [1] and therefore the capacity, throughput, link reliability, spectral efficiency and transmission range. Network coding has been widely used in different types of networks, ranging from Internet to wireless sensor networks, traditional wireless networks, video multicast networks, Peer-to-Peer (P2P) networks and many others [2]. However, limited efforts have been conducted to date to analyze the potential of network coding in vehicular networks.

Vehicular networks are being designed to improve traffic safety and efficiency thanks to the exchange of positioning and basic status information between vehicular nodes. This exchange is based on the periodic transmission and reception of beacons (1-hop broadcast messages) using the IEEE 802.11p technology. To effectively support vehicular safety applications, each vehicle needs to continuously receive updated information from all neighboring vehicles. To this aim, a number of awareness control protocols have been proposed in the literature [3]. Awareness control protocols

are aimed at ensuring each vehicle's capacity to detect, and possibly communicate with the relevant vehicles and infrastructure nodes present in their local neighborhood. All vehicles will periodically transmit their beacons on the so called control channel. This can lead to possible channel congestion, in particular under high traffic densities. The critical nature of the control channel has fostered significant efforts in the research and standardization communities to design congestion control protocols that ensure the scalability and adequate operation of vehicular networks by adapting the transmission parameters of beacons [3]. In fact, the ETSI communications architecture that future connected vehicles will implement includes a key Decentralized Congestion Control (DCC) module that is currently under development [4].

Limited studies have analyzed the use of network coding in vehicular networks. The work in [5] demonstrates that network coding can improve the performance of repetition-based error recovery mechanisms in vehicular networks. With these mechanisms, each vehicle retransmits each beacon k times to recover packets lost due to propagation errors or collisions. The work in [5] proposes that each vehicle XORs its own packet with the packet received from its closest neighbor to improve the probability of successful reception of beacons. Network coding can also be used to improve the performance and efficiency of multi-hop beaconing schemes. The study in [6] analyzes the potential of multi-hop beaconing to improve cooperative awareness in vehicular networks. The theoretical results obtained show that the channel load could be reduced with multi-hop beaconing. However, the conducted simulations showed that packet collisions, the radio channel variability and suboptimal relaying prevent that multi-hop beaconing improves the awareness performance of single-hop beaconing. Finally, the work in [7] was one of the first studies to propose multi-hop beaconing algorithms that exploit network coding to improve cooperative awareness in vehicular networks. In this case, the results obtained showed that the performance (average information age and probability of experiencing a situational-awareness black-out of at least 1 second) can be improved thanks to network coding.

In this context, this paper analyzes the potential benefits of network coding and multi-hop beaconing to reduce the channel load in vehicular networks. While previous studies evaluated network coding schemes based on fixed transmission powers, this paper considers that the

transmission power is adapted to satisfy the application requirements. The first results obtained in this paper show that multi-hop beaconing with network coding could reduce the channel load experienced, which would be important to mitigate the congestion control problem of vehicular networks.

II. BEACONING STRATEGIES

To evaluate the potential of network coding and multi-hop beaconing in vehicular networks, the beaconing strategies depicted in Fig. 1 have been studied. In all of them, all vehicles periodically transmit beacons to support cooperative active safety applications. The applications require that such beacons are successfully received by all vehicles within certain communication range (CR) with certain probability (p_{app}). For example, an application could require that 10 beacons per second are received at a certain communication range (CR) with $p_{app}=0.99$ probability. Both CR and p_{app} are application requirements that depend on the vehicular context. Such requirements can be satisfied with a single-hop (SH) or a multi-hop (MH) strategy. With the multi-hop strategy, each vehicle forwards other vehicles' beacons. For simplicity and based on previous studies [6], only two hops have been considered in this study for the multi-hop beaconing strategies.

With a single-hop beaconing strategy (Fig. 1a), vehicles periodically transmit their beacons using the transmission power needed to reach CR with probability p_{app} . The transmission power for single-hop transmissions (P_t^{SH}) needs to be configured so that the Packet Delivery Ratio (PDR) at a distance equal to CR is p_{app} :

$$p_{app} = PDR(P_t^{SH}, CR) \quad (1)$$

With a multi-hop beaconing strategy (Fig. 1b), vehicles do not only transmit their own beacons, but also forward beacons transmitted by other vehicles. Since all beacons are forwarded by intermediate vehicles, they can be transmitted with lower transmission power levels to reach the target communication range with the desired probability p_{app} . Following the illustration in Fig. 1b, a beacon will be successfully received by vehicle C if either the direct transmission from vehicle A to vehicle C is successful, or if vehicle B successfully receives such transmission from vehicle A and successfully forwards it to vehicle C. While different forwarding strategies are possible, we consider for our analysis that beacons are always forwarded by vehicles located at $CR/2$ distance from the initial transmitter. As a result, the transmission power for multi-hop strategies (P_t^{MH}) needs to satisfy the following equation to satisfy the application requirements:

$$p_{app} = 1 - (1 - PDR(P_t^{MH}, CR)) \cdot (1 - PDR(P_t^{MH}, CR/2))^2 \quad (2)$$

Ec. (2) considers independent transmissions. With the multi-hop beaconing strategy, the transmission power can be reduced to satisfy Ec. (2), but the number of transmissions per vehicle increases. In fact, in a scenario with vehicles uniformly distributed, the average number of packets transmitted per second per vehicle would be 3 times the beacon transmission frequency, because each vehicle would have to forward at least the beacons received from two of its

neighbors (e.g. vehicle B in Fig. 1b would transmit its own beacons plus the beacons received from A and C).

With a multi-hop beaconing strategy combined with network coding (Fig. 1c), beacons can be transmitted with the transmission power that satisfies Ec. (2). However, the beacons forwarded by each vehicle are a combination (XOR operation) of two previously received beacons. In the example shown in Fig. 1c, vehicle B transmits *beacon A+C*, which is the XOR operation of *beacon A* and *beacon C*. Vehicle C can retrieve *beacon A* from *beacon A+C* because it knows *beacon C*. In a scenario with vehicles uniformly distributed, the number of packets to be transmitted per second per vehicle would be 2 times the beacon transmission frequency, because the two beacons to be forwarded are combined with network coding before transmission.

To avoid limiting this study to a specific forwarding strategy (with or without network coding), we will consider a network of vehicles uniformly distributed. Each vehicle will configure its transmission power to achieve the target packet reception probability p_{app} following Ec. (1) or (2). For the analysis of multi-hop beaconing (with and without network coding), vehicles will forward all beacons received from vehicles located at $CR/2$. While more complex situations are obviously possible (e.g. vehicles do not need to be located at $CR/2$ meters), this simplification allows this study to be independent of specific forwarding algorithms, and to quantify in a simple but effective way the potential of network coding and multi-hop beaconing. The fundamental question to be solved is whether a vehicle transmitting a high number of beacons at low power is more efficient than a vehicle transmitting a lower number of beacons, but with higher power. To fairly compare all approaches, all of them need to be configured to satisfy the application requirements.

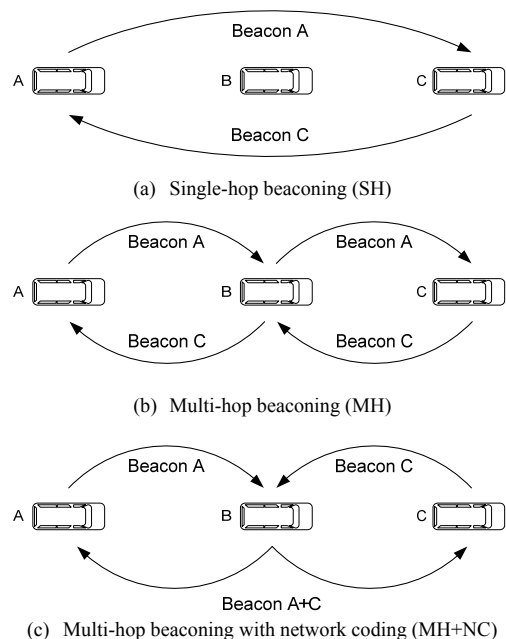


Fig. 1. Beaconing strategies evaluated.

III. EVALUATION

A. Simulation settings

The performance and efficiency of the beaconing strategies analyzed have been evaluated using the network simulator ns-2.35. The simulations conducted consider a straight highway with 6 lanes. In this scenario, vehicles are uniformly distributed and the traffic density considered is 20 vehicles/km/lane.

In this scenario, each vehicle transmits 10 beacons per second using IEEE 802.11p at 6Mbps. The transmission power is adjusted for each simulation to satisfy the application requirements, i.e. to successfully receive all beacons at CR with p_{app} probability in an scenario without interferences. Both CR and p_{app} influence the transmission parameters needed and therefore the channel load generated. To avoid limiting this study to a given application, different combinations of CR and p_{app} have been analyzed.

Taliwal et al. showed in [8] that the Nakagami- m distribution suitably describes the radio propagation conditions in vehicular networks on highways in the absence of interferences. Following [9], this study utilizes the Nakagami- m propagation model with $m=3$ and a quadratic path-loss according to the Friis model.

Table I summarizes the main communication and simulation parameters considered in this study.

Table I
COMMUNICATION AND SIMULATION PARAMETERS

Parameter	Value
Number of lanes	6
Road length [km]	8
Traffic density [veh/km/lane]	20
Packet transmission frequency [Hz]	10
Payload size [Bytes]	250
Data rate [Mbps]	6
Carrier frequency [GHz]	5.9
SINR min for packet reception [dB]	8
Noise floor [dBm]	-99
Communication Range (CR)[m]	100, 200, ..., 500
Probability of successful reception (p_{app})	0.99, 0.999, 0.9999
Simulation time [s]	50
Simulation runs	5

B. Results

To analyze the potential channel load benefits of the use of network coding and multi-hop beaconing, the transmission parameters need to be fairly configured. To this aim, the packet delivery ratio curves for a wide range of transmission power levels have been obtained. Fig. 2 shows as an example, the PDR curves for five different transmission power levels when considering the Nakagami- m propagation model with $m=3$ and quadratic path-loss without interferences. These PDR curves have been used to identify the minimum transmission power level needed to satisfy the application requirements, i.e. to achieve the target probability of packet reception (p_{app}) at the required communication range (CR). Ec. (1) and (2) have been used to find the relationship between PDR and p_{app} for single-hop and multi-hop strategies. The resulting transmission power levels needed to achieve $p_{app}=0.99, 0.999$ and 0.9999 at varying communication ranges are shown in Fig. 3 for both single-hop and multi-hop beaconing strategies. As it can be

observed, the transmission power can be reduced between 6.5dB and 8dB when multi-hop transmissions are considered, while satisfying the application requirement p_{app} at the target CR. The transmission power reduction is higher as the p_{app} parameter increases. For the specific case of $p_{app}=0.99$ and CR=200m, Fig. 4 shows the PDR curves for single-hop and multi-hop beaconing strategies when the transmission power is accurately calculated to satisfy the application requirements. As it can be observed, while a PDR of 0.99 needs to be obtained at CR=200m for single-hop transmissions, the PDR (and therefore the transmission power) can be reduced when multi-hop transmissions are allowed.

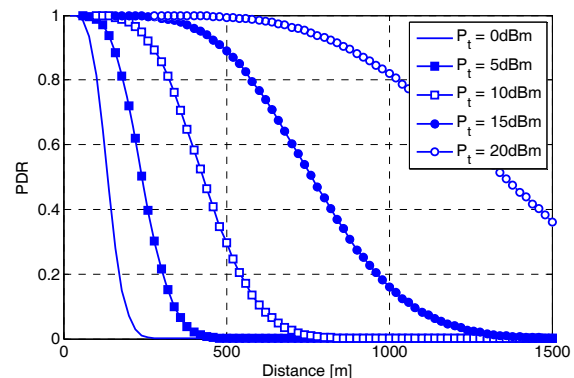


Fig. 2. PDR (Packet Delivery Ratio) for different transmission power levels considering Nakagami- m propagation model with $m=3$ and quadratic path-loss without interferences.

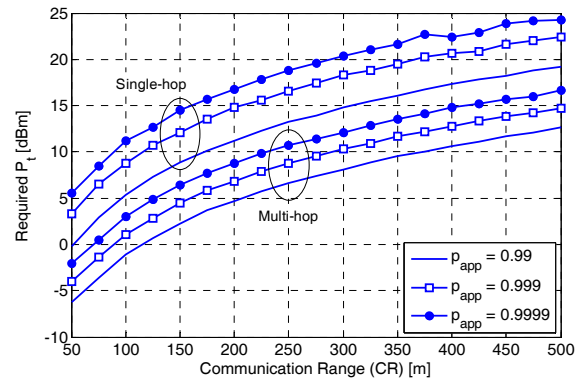


Fig. 3. Required transmission power to obtain a given probability of successful reception (p_{app}) at a certain communications range (CR) without interferences.

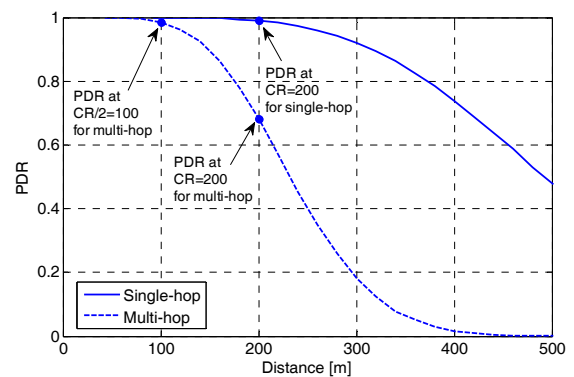


Fig. 4. PDR (Packet Delivery Ratio) curves for the transmission power levels needed to achieve $p_{app}=0.99$ at CR=200m for single-hop and multi-hop strategies.

To evaluate the channel load benefits of network coding and multi-hop beaconing, this study considers that all vehicles in the scenario have the same application requirements (p_{app} and CR), and configure their transmission power to satisfy them. Therefore, for the single-hop strategy, all vehicles are configured to transmit 10 beacons per second using the transmission power level shown in Fig. 3 for the p_{app} and CR parameters considered. For multi-hop strategies, we will consider that each vehicle forwards all beacons received from vehicles located at $CR/2$ on the same lane. As a result, when considering a multi-hop strategy without network coding, each vehicle will forward 20 packets per second (10 from a vehicle located $CR/2$ meters ahead, and 10 from a vehicle located at $CR/2$ behind). When considering network coding, only 10 packets per second need to be forwarded. This is the case because each beacon from the vehicle located at $CR/2$ meters ahead is combined (XOR) with the beacon received from the vehicle located at $CR/2$ meters behind. As a consequence, in average, all vehicles transmit 10Hz with the single-hop beaconing strategy, 30Hz with the multi-hop beaconing strategy without network coding, and 20Hz with the multi-hop beaconing strategy with network coding.

The trade-off between transmission power and packet transmission frequency for the different strategies is illustrated in Fig. 5. This figure represents the number of packets that would be sensed per second by a vehicle located at a certain distance from a given transmitter. The transmitter is configured with the transmission power needed to achieve $p_{app}=0.99$ at $CR=200m$ and the number of packets transmitted per second include the packets that would be forwarded if a multi-hop strategy was considered. The channel load experienced is directly related to the time the radio interface is busy sensing packets sent by other radio interfaces. A multi-hop beaconing strategy increases the number of packets sensed per second at short distances, but decreases it at medium and high distances.

Fig. 6 shows the channel busy ratio (CBR) experienced by a vehicle located in the center of the scenario for different beaconing strategies and varying application requirements (CR and p_{app}). The CBR represents the percentage of time that the radio interface of a vehicle is busy and is widely used as a channel load metric. As it can be observed, the CBR increases as the application requirements (CR and p_{app})

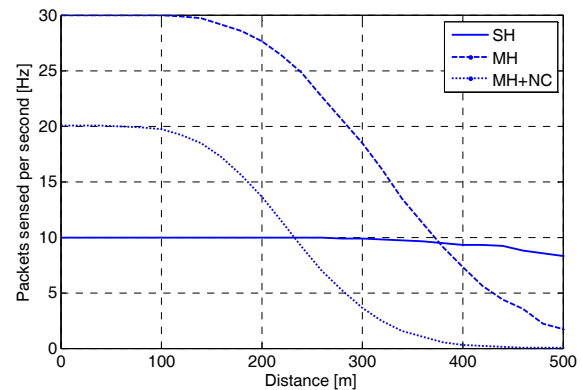


Fig. 5. Number of packets sensed per second as a function of the distance to a transmitter considering the transmission power levels needed to achieve $p_{app}=0.99$ at $CR=200m$ for single-hop and multi-hop strategies (with and without network coding).

increase. This is the case because the increase of the application requirements augments the transmission power needed to satisfy them in absence of interference, as previously shown in Fig. 3. As it can be observed, the channel starts saturating for CBR values higher than 0.8, and therefore further increasing the CR does not significantly increase the channel load.

To more clearly compare the CBR experienced with the different beaconing strategies, the relative variation (RV) of the CBR with respect to the single-hop strategy has been calculated. RV has been calculated with the following equation:

$$RV = 100 \cdot \frac{CBR - CBR_{SH}}{CBR_{SH}} \quad (3)$$

where CBR_{SH} is the CBR experienced with the single-hop strategy. While a positive RV value represents a channel load increase compared to the single-hop beaconing strategy, a negative value represents a channel load reduction. Fig. 7 depicts the relative variation of the CBR with respect to the single-hop strategy for varying application requirements. As it can be observed, the channel load can increase up to around 25% with multi-hop beaconing (MH) compared with the single-hop beaconing strategy. However, when multi-hop beaconing is combined with network coding (MH+NC), the channel load can be reduced. As shown in Fig. 7, the channel

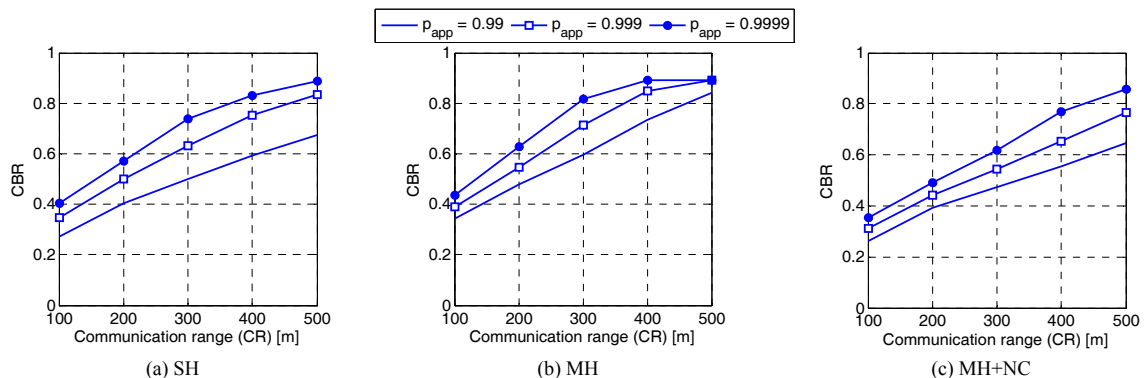


Fig. 6. CBR (Channel Busy Ratio) experienced by a vehicle in the center of the scenario for single-hop (SH), multi-hop (MH) and multi-hop combined with network coding (MH+NC) strategies. All vehicles are configured with the transmission parameters needed to satisfy the application requirements (CR and p_{app}) in absence of interference.

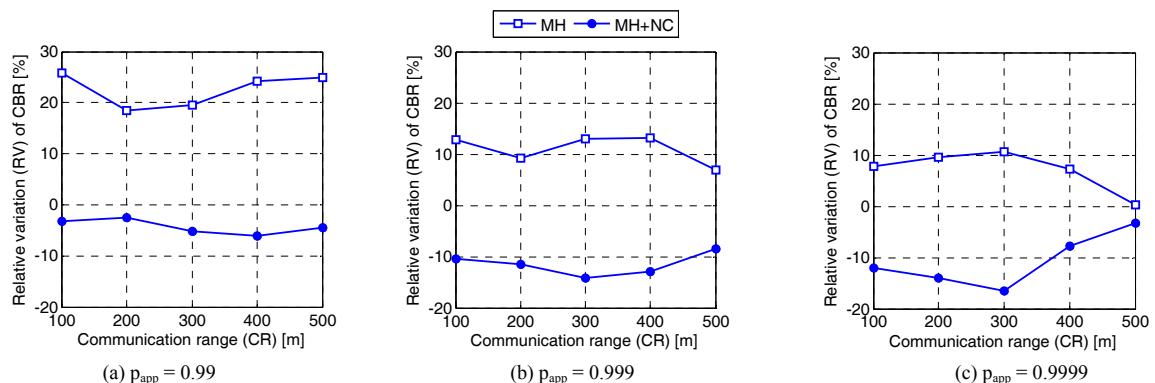


Fig. 7. Relative variation of the CBR (Channel Busy Ratio) experienced by a vehicle in the center of the scenario for multi-hop (MH) and multi-hop combined with network coding (MH+NC) strategies. The values shown represent the relative variation with respect to single-op beaconing (SH), i.e. a positive (negative) value means an increase (decrease) of the channel load generated compared to SH.

load can be reduced with the strategy that combines multi-hop beaconing with network coding up to around 17%. This reduction is especially relevant given the critical nature of the control channel used for vehicular networks.

Interestingly, the performance and efficiency gain achieved with network coding is independent of certain communication parameters. In particular, it is independent of the packet payload and data rate. Both parameters affect the time duration of a packet, and therefore affect the channel load generated. However, the relative gain that can be achieved with network coding in terms of channel load is maintained irrespective of the specific values considered. The results have not included in this paper due to space limitations and small differences with the results presented.

IV. DISCUSSION

First results presented in this paper show that network coding and multi-hop beaconing can reduce the channel load of vehicular networks. The reduction of the transmission power and the combination of packets to be forwarded reduce the channel load in the scenario considered and therefore decrease the interferences generated. Further analysis will be needed to verify if the same conclusions can be obtained in different scenarios. The potential reduction of the channel load can be especially relevant for the vehicular control channel, which can be easily congested especially under high traffic density conditions. In fact, the fundamental limits of solutions based on IEEE 802.11p can be reached when the application requirements and traffic density are

high [10]. If the benefits of network coding are verified for different scenarios and conditions, it could be part of a potential extension of the DCC algorithms being discussed at ETSI to ensure the scalability of vehicular networks, which will be especially challenging with the emergence of automated vehicles. Future automated vehicles will require the wireless exchange of richer information and higher traffic densities at higher speeds might be possible, which would notably increase the channel load and communication requirements.

In scenarios with large obstacles, such as buildings or trucks, the use of multi-hop beaconing could represent the only way to reach the target communication range with the required probability due to the power limitations. Large obstacles produce high propagation losses, especially due to the high carrier frequency of IEEE 802.11p [11][12]. As a result, the transmission power needed to satisfy the application requirements could be higher than the limits imposed by the standards (33dBm). Multi-hop beaconing can reduce the transmission power levels needed compared to single-hop beaconing. This reduction can be especially significant when single-hop transmissions are blocked by large obstacles, but multi-hop transmissions are produced under LOS (Line-of-Sight) conditions (see examples in Fig. 8). Since multi-hop beaconing can be very inefficient, as previously demonstrated in Fig. 7, network coding can be used to improve its efficiency, as demonstrated in the results obtained in section III.

To exploit the potential of network coding, an efficient forwarding algorithm that is compliant with the DCC

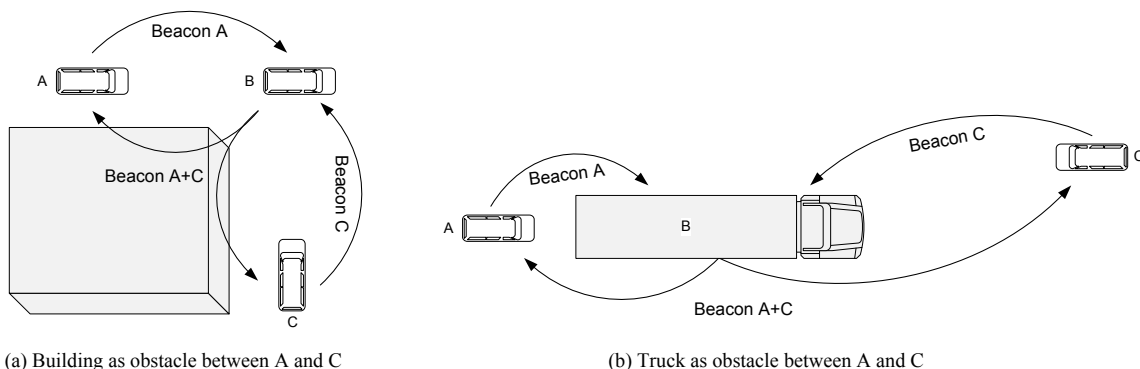


Fig. 8. Scenarios that are especially suitable for multi-hop beaconing and network coding.

architecture discussed at ETSI needs to be designed. The forwarding algorithm needs to dynamically select the neighboring vehicles whose beacons will be combined with network coding and transmitted [7]. This is one of the key elements of the algorithm to maximize the number of vehicles that are able to decode the forwarded beacon, while satisfying the application requirements. Moreover, a reliable and efficient power control algorithm needs to be designed to achieve the channel load gains shown in this paper. In fact, the application requirements can be dynamic and context dependent [13], which challenge the accurate adaptation of the transmission power. The forwarding algorithm should also be able to ensure that the application requirements are satisfied.

V. CONCLUSIONS

The results obtained in this paper show that network coding can reduce the channel load generated by the beaconing process of vehicular networks. The channel load generated when considering a multi-hop beaconing strategy combined with network coding can be up to around 17% lower than the channel load generated with a single-hop beaconing strategy. Both strategies have been fairly compared by adapting the transmission power to the minimum needed to satisfy the application requirements in absence of interference. The results obtained will need to be confirmed for different scenarios and conditions, but could motivate the design of efficient multi-hop beaconing algorithms that exploit network coding to reduce the channel load. These algorithms could be relevant for the future evolution of the DCC architecture being discussed at ETSI, especially due to the emergence of automated vehicles which will require the transmission of richer information. Further work will be needed to analyze the effect of multi-hop beaconing and network coding on packet collisions and interferences, which will especially depend on the forwarding algorithm.

ACKNOWLEDGEMENT

This work was supported in part by the Spanish Ministry of Economy and Competitiveness and FEDER funds TEC2014-57146-R and by the Spanish *Dirección General de Tráfico* (SPIP20141414).

REFERENCES

- [1] P.A. Chou, Y. Wu, "Network Coding for the Internet and Wireless Networks", *IEEE Signal Processing Magazine*, vol.24, no.5, pp. 77-85, Sept. 2007.
- [2] M. Z. Farooqi et al., "A survey on network coding: From traditional wireless networks to emerging cognitive radio networks", *Journal of Network and Computer Applications*, vol. 46, 166-181, 2014.
- [3] M. Sepulcre et al., "Congestion and Awareness Control in Cooperative Vehicular Systems", *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1260-1279, July 2011.
- [4] ETSI TC ITS, "Intelligent Transport Systems (ITS); Cross Layer DCC Management Entity for operation in the ITS G5A and ITS G5B medium; Report on Cross layer DCC algorithms and performance evaluation", *ETSI TR 101 612 V1.1.1*, Sept. 2014.
- [5] Z. Wang; M. Hassan, "Network Coded Repetition: A Method to Recover Lost Packets in Vehicular Communications", *Proc. IEEE International Conference on Communications (ICC)*, Kyoto, Japan, pp. 1-6, 5-9 June 2011.
- [6] J. Mittag, F. Thomas, J. Harri, H. Hartenstein, "A Comparison of Single- and Multi-Hop Beaconing in VANETs", *Proc. ACM International Workshop on Vehicular Inter-NETworking (VANET)*, Beijing, China, pp. 69-78, 25 Sept. 2009.
- [7] F. Librino, M. E. Renda, P. Santi, "Evaluating Multi-Hop Beaconing Forwarding Strategies for IEEE 802.11p Vehicular Networks", *Proc. IEEE Vehicular Networking Conference (VNC)*, Boston, USA, pp. 31-38, 16-18 Dec. 2013.
- [8] V Taliwal, D Jiang, H Mangold, C Chen, R Sengupta, "Empirical determination of channel characteristics for DSRC vehicle-to-vehicle communication", *Proc. ACM International Workshop on Vehicular Ad Hoc Networks (VANET)*, Philadelphia, USA, October 2004.
- [9] M. Killat, H. Hartenstein, "An empirical model for probability of packet reception in vehicular ad hoc networks", *EURASIP Journal on Wireless Communications and Networking*, Jan. 2009.
- [10] M. Sepulcre, J. Gozalvez, O. Altintas, H. Kremling, "Integration of Congestion and Awareness Control in Vehicular Networks", [Submitted to Ad Hoc Networks].
- [11] M. Sepulcre and J. Gozalvez, "Experimental evaluation of cooperative active safety applications based on V2V communications", *Proc. ACM international workshop on Vehicular inter-networking, systems, and applications (VANET)*, Low Wood Bay (United Kingdom), pp. 13-20, 25 June 2012.
- [12] J. Gozalvez, M. Sepulcre and R. Bauza, "IEEE 802.11p Vehicle to Infrastructure Communications in Urban Environments", *IEEE Communications Magazine*, vol. 50, no. 5, pp. 176-183, May 2012.
- [13] M. Sepulcre and J. Gozalvez, "On the importance of application requirements in cooperative vehicular communications", *Proc. International Conference on Wireless On-Demand Network Systems and Services (WONS)*, pp.124-131, Bardonecchia (Italy), 26-28 January 2011.

Downlink Optimisation of FFR-Aided OFDMA Networks using Channel-Aware Scheduling

Jan García-Morales, Guillem Femenias, and Felip Riera-Palou

Mobile Communications Group – University of the Balearic Islands (UIB) – Palma 07122, Illes Balears (Spain)

Email: {jan.garcia,guillem.femenias,felip.riera}@uib.es

Abstract—Modern cellular standards typically incorporate interference coordination schemes allowing near universal frequency reuse while preserving reasonably high spectral efficiencies over the whole coverage area. In particular, fractional frequency reuse (FFR) and its variants are deemed to play a fundamental role in the next generation of cellular deployments (B4G/5G systems). This paper presents an analytical framework allowing the downlink performance evaluation of FFR-aided OFDMA networks when using channel-aware scheduling policies. Analytical performance results are used to optimise the FFR parameters as a function of, for instance, the resource block scheduling policy or the density of UEs per cell. Furthermore, different optimisation designs of the FFR component are proposed that allow a tradeoff between throughput performance and fairness by suitably dimensioning the FFR inner and outer areas and the corresponding frequency allocation to each region.

I. INTRODUCTION

Orthogonal frequency division multiple access (OFDMA) has been adopted as the downlink multiple access scheme for state-of-the-art cellular communications standards such as the Long-Term Evolution (LTE) and LTE-Advanced (LTE-A) [1]. In OFDMA networks, a wideband frequency-selective fading channel is decomposed into a set of orthogonal narrowband subchannels. These subchannels are jointly used with a time-slotted frame structure to provide a set of frequency/time resources (also known as resource blocks (RBs)), which are distributed among cells based on predefined frequency reuse plans. The orthogonality among RBs makes the intra-cell interference negligible. However, the use of *aggressive* high spectral efficiency universal frequency reuse plans, with all cells using the same set of RBs, cause the OFDMA-based networks to suffer from very high levels of inter-cell interference (ICI), particularly affecting the UEs located in the cell-edge areas. With the aim of mitigating ICI experienced by the cell-edge users while still achieving high spectral efficiencies, a myriad of ICI control (ICIC) strategies have been proposed [2], among which *static* fractional frequency reuse (FFR) and all its variants show a good tradeoff between cell-edge throughput enhancement, provision of high spectral efficiency and implementation complexity [3].

In the downlink of FFR-aided OFDMA-based networks, the time/frequency channel quality of the set of RBs allocated to a given base station (BS), typically measured in terms of the signal-to-interference-plus-noise ratio (SINR), varies for different UEs. Such variations in channel conditions can be exploited by using channel-aware schedulers able to allocate each RB to a UE with favourable channel conditions at a given scheduling time slot. Opportunistic maximum SINR (MSINR) schedulers [4] makes the most of the multiuser diversity by

allocating the RBs to the UEs experiencing the best channel conditions. Although applying this scheduling rule aims at maximising the spectral efficiency of the system, it raises a serious fairness problem, with UEs suffering from bad channel conditions over extended periods of time experiencing a dramatic quality-of-service (QoS) degradation. In order to provide a reasonable tradeoff between spectral efficiency and fairness, a proportional fair (PF) scheduling rule was proposed by Kelly *et al.* in [5] and then extended by Shakkottai *et al.* in [6]. In this case, scheduling decisions are based on a weighted version of the instantaneous channel behaviour, with the weighting coefficient for a given UE being inversely proportional to the average channel behaviour during the time this UE has gained access for transmission. In this way, the RBs end up being allocated to UEs experiencing the relatively best channel conditions in comparison to their average channel state and thus, using a PF scheduler, the possibility of a UE with a very bad link suffering from long periods of starvation is drastically reduced.

The problem of analytical performance evaluation of FFR-aided OFDMA-based cellular networks has been recently tackled using tools drawn from the theory of stochastic geometry, where the BSs are distributed using Poisson Point Processes (PPPs) [7] (see also [8] and references therein). Stochastic geometry-based frameworks analyze the system performance by spatially averaging over all possible network realisations. Using this approach allows characterising the performance for an entire network but precludes from accurately analysing the performance of a given cell, a metric of particular importance to network designers that, provided a planned set of BS locations along with traffic load conditions, may be interested in calculating the performance obtained within a specific region in the coverage area of the network. In order to overcome the limitations of the stochastic geometry approach, Heath *et al.* [9] proposed a rather involved upper bound to model the interference by inscribing a circular cell within the weighted Voronoi cell provided by the PPP-based model. Nonetheless, characterising the cellular layouts using stochastic geometry makes it extremely difficult, if at all possible, to accurately model the use of powerful ICIC techniques proposed for current and next-generation well-planned macrocell networks, including FFR and its variants and/or BS cooperation schemes.

In contrast to the above background work, which relies on the use of stochastic geometry to model the cellular environment, Fan Jin *et al.* [10] considered an FFR-aided two-layer OFDMA network where stochastic geometry was used to characterise the random distribution of femtocells, and the

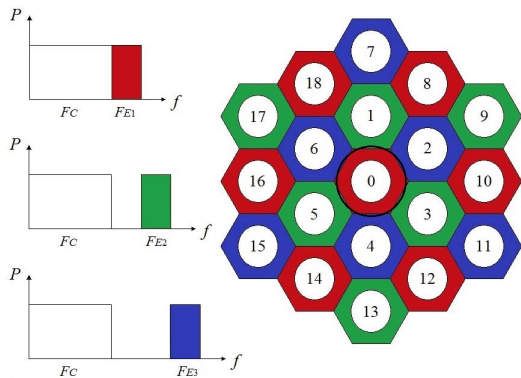


Fig. 1: Topology of the FFR-aided OFDMA-based two-tier network.

macrocells were overlaid on top of the femtocells following a regular tessellation. One of the main limitations of this work, however, was the use of rather unrealistic assumptions such as neglecting the small scale fading effects and, consequently, limiting the proposed analytical framework to resource allocation schemes based on the round robin scheduling policy. Similar approaches, lacking the consideration of small scale fading and scheduling policies, were also proposed by Assaad in [11] and Najjar *et al.* in [12] to optimise FFR-based parameters in a single-tier network. These limitations were overcome in part in our recent contribution [13] (see also [14]), but only taking into account the use of opportunistic MSINR schedulers.

Despite the attractive properties of the PF scheduler, it makes the spectral efficiency analysis of the system much more complicated [15]–[18]. In this paper we present a novel analytical framework allowing the throughput performance evaluation of an FFR-aided OFDMA-based cellular network using a PF scheduling policy. The proposed analytical framework turns out to extend the applicability of the analysis to the optimisation of the FFR in the spatial and frequency domains using different optimisation criteria, namely, the fixed-spectrum-allocation-factor design, the area-proportional design or the QoS-constrained design. At the same time, it is shown how the performance of both the round robin (RR) and the MSINR scheduling rules can be derived as special cases of the PF scheduler. Although results are obtained for an FFR-aided deployment, this analytical framework opens the door to the theoretical spectral efficiency evaluation of OFDMA-based cellular networks using more sophisticated ICIC techniques such as soft frequency reuse, adaptive frequency reuse or network MIMO, as well as to the assessment of cellular multi-tier networks where the macro-cellular network is underlaid by different tiers of pico- and femto-cellular BSs.

II. SYSTEM MODEL

A. Network topology model

Let us consider the downlink of an FFR-aided OFDMA-based (LTE-like) cellular system where a set of macrocellular BSs are assumed to be deployed following a conscious planning and thus, are regularly arranged over the whole coverage

area. This cellular environment can be safely modeled as a regular tessellation of hexagonally-shaped coverage areas, as shown in Fig. 1, with the BSs located at the centre of the hexagons¹. For the sake of analytical tractability, the central cell, covered by BS 0, which will be referred to as the tagged BS or the BS of interest, will be approximated by a circle whose area is the same as the hexagonal one (see Fig. 1). That is, assuming that the side of the regular hexagon is R_h , the radius of the circular cell is $R_m = R_h \sqrt{3\sqrt{3}/(2\pi)}$.

In order to control the ICI, UEs are classified according to the received average signal-to-interference-plus-noise ratio (SINR) as either cell-centre users, when the received SINR is above a given threshold, or cell-edge users, when it is below the threshold. An FFR scheme is then applied by allocating non-overlapping frequency bands to cell-centre and cell-edge users, while employing a frequency reuse factor equal to one for the cell-centre users and a higher frequency reuse factor for the cell-edge users that, without loss of generality, is assumed to be three in this paper. Again, for analytical tractability, cell-centre and cell-edge regions will be separated by a circumference of radius R_{th} (threshold distance or threshold radius). The total system bandwidth is exploited by means of a set \mathcal{F}_T of subcarriers, which is split into a set \mathcal{F}_C of subcarriers allocated to the cell-centre and a set $\mathcal{F}_T \setminus \mathcal{F}_C$ of subcarriers allocated to the cell-edge. The set $\mathcal{F}_T \setminus \mathcal{F}_C$ is further split into three equal parts, namely \mathcal{F}_{E1} , \mathcal{F}_{E2} and \mathcal{F}_{E3} , of size $|\mathcal{F}_E|$, which are allocated to cell-edge UEs in such a way that adjacent cells will operate on different sets of subcarriers, as shown in Fig. 1.

The locations of the UEs at a given time instant are assumed to form a stationary PPP of normalised intensity λ_m (measured in UEs per area unit). The PPP assumption is valid when the users are independently and uniformly distributed over the service coverage area, which is a reasonable approach to obtain analytical expressions for the spectral efficiency of the network. A consequence of this assumption is that the probability distribution of the number M_S of users falling within any spatial region \mathcal{S} of area A_S follows a Poisson distribution, thus implying

$$Pr\{M_S = M\} = \frac{(\lambda_m A_S)^M e^{-\lambda_m A_S}}{M!}. \quad (1)$$

B. Channel model

The downlink channel is subject to path loss and small-scale fading². The path loss characterising the link between the b th BS and the u th UE can be modeled as (in dB)

$$L(d_{b,u}) = K + 10\alpha \log_{10}(d_{b,u}), \quad (2)$$

where K and α are, respectively, a fixed path loss and the path loss exponent, and $d_{b,u}$ is the distance (in metres) between the UE and the BS.

¹Omnidirectional antenna BSs are assumed in this paper. In future work this will be extended to consider the use of sectorisation.

²In line with Xu *et al.* [19] and Jin *et al.* [10], note that, for the sake of analytical simplicity, only pathloss and small scale fading are considered in this paper. In future work this will be extended to consider large scale fading (shadowing) as well.

The instantaneous SINR experienced by UE u in the cell of interest on the n th subcarrier during the scheduling period t can be expressed as

$$\gamma_{u,n}(t) = \frac{P_c L(d_{0,u}) |H_{0,u,n}(t)|^2}{N_0 \Delta f + I_{u,n}(t)}, \quad (3)$$

where, assuming the use of uniform power allocation, $P_c = P_T / (|\mathcal{F}_C| + |\mathcal{F}_E|)$ is the power allocated per subcarrier, with P_T denoting the available transmit power at the BS, $H_{b,u,n}(t) \sim \mathcal{CN}(0,1)$ is the frequency response resulting from the small-scale fading channel linking the b th BS to UE u on the n th subcarrier during scheduling period t , N_0 is the noise power spectral density, and $I_{u,n}(t)$ denotes the interference term that is given by

$$I_{u,n}(t) = \sum_{b \in \Phi_n} P_c L(d_{b,u}) |H_{b,u,n}(t)|^2, \quad (4)$$

with Φ_n representing the set of interfering BSs, which is subcarrier-dependent as, because of the use of FFR, the set of interfering BSs for subcarriers allocated to the cell-centre is different from that for subcarriers allocated to the cell-edge. As an important notational remark, note that $L(d_{b,u})$ can be expressed in terms of UEs' polar positions with respect to BS 0 as $L(d_{0,u}, \theta_{0,u})$ and thus, strictly speaking, $\gamma_{u,n}(t)$ is a function of $d_{0,u}$ and $\theta_{0,u}$. Nevertheless, it is shown in [20] that the instantaneous SINR in multi-cell networks barely depends on the polar angle and thus, from this point onwards, the dependence of $\gamma_{u,n}(t)$ on $\theta_{0,u}$ will be omitted.

III. THROUGHPUT ANALYSIS

Let us define M_0 as a positive integer random variable representing the number of UEs in the region served by the tagged BS. Using this definition, the average cell throughput for the downlink of the FFR-aided OFDMA-based cellular network can be expressed as

$$\eta = \sum_{k=1}^{\infty} Pr\{M_0 = k\} \sum_{k_C=0}^k \binom{k}{k_C} P_C^{k_C} (1 - P_C)^{k-k_C} \times [|\mathcal{F}_C| \eta_n^C(k_C) + |\mathcal{F}_E| \eta_n^E(k - k_C)], \quad (5)$$

where P_C denotes the probability that a UE is located in the cell-centre region and $\eta_n^A(k)$ is the average cell throughput on the n th subcarrier when there are k UEs in cell area A , where A is a token used to represent either the cell-centre region C or the cell-edge region E .

As UEs are assumed to be uniformly distributed in the entire cell region, the probability that a UE is located in the cell-centre area is

$$P_C = \frac{R_{th}^2 - R_{0m}^2}{R_n^2 - R_{0m}^2}, \quad (6)$$

where R_{0m} denotes the minimum distance of a UE from its serving BS. Now, defining M_A as a non-negative integer random variable representing the number of UEs in the cell region A , the average cell throughput on the n th subcarrier allocated to cell region A when $M_A = k$, can be obtained as

$$\begin{aligned} \eta_n^A(k) &= \mathbb{E}_{\gamma_n^A | M_A} \{ \Delta f \log_2(1 + \gamma_n^A) | M_A = k \} \\ &= \Delta f \log_2 e \int_0^{\infty} \frac{1 - F_{\gamma_n^A | M_A}(x|k)}{1+x} dx, \end{aligned} \quad (7)$$

where γ_n^A denotes the instantaneous SINR experienced on the n th subcarrier of cell area A , $F_{\gamma_n^A | M_A}(x|k)$ is its cumulative distribution function (CDF) and Δf denotes the subcarrier bandwidth.

In order to obtain closed-form average throughput expressions, the CDF $F_{\gamma_n^A | M_A}(x|k)$ has to be calculated and this depends on the specific scheduling policy applied by the resource allocation algorithm. In the following subsections, this CDF will be obtained for the PF scheduling rule and then it will be extended to both the MSINR and the RR schedulers.

A. PF scheduling

As stated by Choi *et al.* in [15], a proportional fair (PF) scheduler, exploiting the knowledge of the instantaneous SINRs experienced by all UEs $q \in \mathcal{M}_A$, allocates subcarrier $n \in \mathcal{F}_A$ to UE $u \in \mathcal{M}_A$ satisfying

$$u = \arg \max_{q \in \mathcal{M}_A} \{ w_q(t) \gamma_{q,n}(t) \}, \quad (8)$$

where \mathcal{M}_A is the set indexing all UEs in cell region A , and $w_q(t) = 1/\mu_q(t)$ is the weighting (prioritization) coefficient for UE q that, in this case, depends on the short-term averaged evolution of channel-state information, which can be obtained using a moving average over a window of W scheduling periods as

$$\mu_q(t) = \left(1 - \frac{1}{W} \right) \mu_q(t-1) + \sum_{n \in \mathcal{F}_A} \iota_{q,n}(t) \frac{\gamma_{q,n}(t)}{W}, \quad (9)$$

with $\iota_{q,n}(t)$ denoting the indicator function of the event that UE q is scheduled to transmit on subcarrier n during scheduling period t , that is,

$$\iota_{q,n}(t) = \begin{cases} 1, & \text{if UE } q \text{ is scheduled on carrier } n \text{ in slot } t \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

From (9) we know that, for large values of W and once the PF scheduler reaches stability, $\mu_q(t)$ varies very little with t and thus, it can be safely approximated by its statistical expectation, that is, $\mu_q(t) \simeq \mathbb{E}\{\mu_q(t)\} \triangleq \bar{\mu}_q$. In fact, as stated by Liu and Leung in [16], experiments suggest that this approximation is valid for $W \geq 50$ with an accuracy greater than 98%. Hence, using this approximation and according to the previous definition of the PF scheduler, UE $u \in \mathcal{M}_A$ will be scheduled on subcarrier $n \in \mathcal{F}_A$ whenever

$$\varphi_{u,n}(t) > \varphi_{\max,u,n}(t) \triangleq \max_{q \in \mathcal{M}_A, q \neq u} \{ \varphi_{q,n}(t) \}, \quad (11)$$

where $\varphi_{q,n}(t) \triangleq \gamma_{q,n}(t) / \bar{\mu}_q$. That is, UE $u \in \mathcal{M}_A$ is allocated subcarrier n during time slot t if

$$\gamma_{u,n}(t) > \bar{\mu}_u \varphi_{\max,u,n}(t). \quad (12)$$

Thus, taking into account that the random variables $\{\varphi_{q,n}(t)\}_{q \in \mathcal{M}_A}$ are independent, the conditional CDF of γ_n^A , conditioned on the event that there are $M_A = k$ UEs

in region A and on the set of distances $\mathbf{d} = \{d_{0,u}\}_{\forall u \in \mathcal{M}_A}$, can be readily evaluated as

$$\begin{aligned}
 F_{\gamma_n^A | M_A, \mathbf{d}}(\gamma | k, \mathbf{d}) &= \Pr \{ \gamma_n^A \leq \gamma | M_A = k, \mathbf{d} \} \\
 &= \sum_{\forall u \in \mathcal{M}_A} \Pr \left\{ \gamma_{u,n}(t) \leq \gamma, \varphi_{\max, u, n}(t) \leq \frac{\gamma_{u,n}(t)}{\bar{\mu}_u} \mid \mathbf{d} \right\} \\
 &= \sum_{\forall u \in \mathcal{M}_A} \int_0^\gamma f_{\gamma_{u,n} | d_{0,u}}(x | d_{0,u}) F_{\varphi_{\max, u, n} | \mathbf{d}} \left(\frac{x}{\bar{\mu}_u} \mid \mathbf{d} \right) dx \\
 &= \sum_{\forall u \in \mathcal{M}_A} \int_0^\gamma f_{\gamma_{u,n} | d_{0,u}}(x | d_{0,u}) \\
 &\quad \times \prod_{\substack{q \in \mathcal{M}_A \\ q \neq u}} F_{\varphi_{q,n} | d_{0,q}} \left(\frac{x}{\bar{\mu}_u} \mid d_{0,q} \right) dx, \tag{13}
 \end{aligned}$$

where $F_{\varphi_{\max, q, n} | \mathbf{d}}(x | \mathbf{d})$ is the conditional CDF of $\varphi_{\max, q, n}(t)$ conditioned on the set of distances \mathbf{d} , and $f_{\gamma_{q,n} | d_{0,q}}(x | d_{0,q})$ and $F_{\varphi_{q,n} | d_{0,q}}(x | d_{0,q})$ are used to denote, respectively, the conditional PDF of $\gamma_{q,n}(t)$ and the conditional CDF of $\varphi_{q,n}(t)$ conditioned on $d_{0,q}$.

In order to obtain further analytical simplifications, let us assume that on each subcarrier n , the conditional random variables $\{\varphi_{q,n} | d_{0,q}\}_{\forall q \in \mathcal{M}_A}$ are independent and identically distributed (i.i.d.). That is, given the positions of UEs in region A , it is assumed that on each subcarrier n in region A the UEs are statistically equivalent in terms of the scheduling metrics. As stated by Jingxian Wu *et al.* in [17], this assumption makes intuitive sense because the fairness of the PF scheduling rule comes indeed from the fact that the scheduling metrics are approximately identically distributed. Using this general assumption and applying integration by parts, the conditional CDF in (13) simplifies to

$$F_{\gamma_n^A | M_A, \mathbf{d}}(x | k, \mathbf{d}) = \frac{1}{k} \sum_{u \in \mathcal{M}_A} F_{\gamma_{u,n} | d_{0,u}}^k(x | d_{0,u}). \tag{14}$$

Now, taking into account that on each subcarrier n in region A , and after averaging over the distance to the BS, the UEs are statistically equivalent in terms of SINR, the (unconditional) random variables $\{\gamma_{q,n}(t)\}_{\forall q \in \mathcal{M}_A}$ are i.i.d., and the conditional CDF in (7) can be obtained as

$$F_{\gamma_n^A | M_A}(x | k) = \int_{R_L^A}^{R_U^A} F_{\gamma_{u,n} | d_{0,u}}^k(x | d) f_{d_{0,u}}(d) dd, \tag{15}$$

where R_L^A and R_U^A denote the lower and upper radiuses of the circumferences defining cell-region A and $f_{d_{0,u}}(d)$ is the PDF of the random variable $d_{0,u}$ that can be expressed as

$$f_{d_{0,u}}(d) = \frac{2d}{R_U^A{}^2 - R_L^A{}^2}, \quad R_L^A \leq d \leq R_U^A. \tag{16}$$

The conditional CDF of the instantaneous SINR $F_{\gamma_{u,n} | d_{0,u}}(x | d)$ in (15) can be derived as [19]

$$\begin{aligned}
 F_{\gamma_{u,n} | d_{0,u}}(x | d) &= 1 - e^{-\frac{x N_0 \Delta f}{\gamma_{0,u}}} \\
 &\quad + \sum_{b \in \Phi_n} \frac{x \bar{\gamma}_{b,u}^{|\Phi_n|}}{(x \bar{\gamma}_{b,u} + \bar{\gamma}_{0,u}) \prod_{\substack{b' \in \Phi_n \\ b' \neq b}} (\bar{\gamma}_{b,u} - \bar{\gamma}_{b',u})} e^{-\frac{x N_0 \Delta f}{\gamma_{0,u}}}. \tag{17}
 \end{aligned}$$

where $\bar{\gamma}_{b,u} = P_c L(d_{b,u})$ is the average received power at UE u from BS b .

B. MSINR scheduling

When implementing the MSINR scheduling rule, in each scheduling period and on each subcarrier n in region A , the BS serves the UE experiencing the highest instantaneous SINR, that is,

$$\gamma_n^A = \max_{q \in \mathcal{M}_A} \{ \gamma_{q,n}(t) \}. \tag{18}$$

Note that the MSINR scheduling rule is equivalent to the scheduler specified in (8) by setting the weighting (prioritization) coefficients to $w_q(t) = 1 \forall q \in \mathcal{M}_A$. In this case, following a reasoning similar to that used to analyse the PF scheduling rule, the conditional CDF of γ_n^A , conditioned on the event that there are $M_A = k$ UEs in region A and on the set $\mathbf{d} = \{d_{0,u}\}_{\forall u \in \mathcal{M}_A}$, can be expressed as

$$F_{\gamma_n^A | M_A, \mathbf{d}}(x | k, \mathbf{d}) = \prod_{u \in \mathcal{M}_A} F_{\gamma_{u,n} | d_{0,u}}(x | d_{0,u}). \tag{19}$$

Consequently, as on each subcarrier n in region A the UEs are statistically equivalent in terms of SINR, the conditional CDF in (7) simplifies to

$$F_{\gamma_n^A | M_A}^{\text{MSINR}}(x | k) = \left[\int_{R_L^A}^{R_U^A} F_{\gamma_{u,n} | d_{0,u}}(x | d) f_{d_{0,u}}(d) dd \right]^k. \tag{20}$$

C. RR scheduling

A RR scheduler allocates subcarriers to UEs in a fair time-sharing approach. Since the SINRs experienced by UEs in region A on each subcarrier n are statistically equivalent, serving $M_A = k$ UEs using a RR scheduling policy is equivalent to serving $M_A = 1$ UEs with MSINR (even when UEs are selected with non uniform probability). Therefore, the conditional CDF in (7) simplifies to

$$\begin{aligned}
 F_{\gamma_n^A | M_A}^{\text{RR}}(x | k) &= F_{\gamma_n^A | M_A}^{\text{MSINR}}(x | 1) \\
 &= \int_{R_L^A}^{R_U^A} F_{\gamma_{u,n} | d_{0,u}}(x | d) f_{d_{0,u}}(d) dd. \tag{21}
 \end{aligned}$$

IV. OPTIMAL DESIGNS

In this section we propose different optimisation designs for the FFR-aided OFDMA-based cellular network aiming at determining the size of the FFR-related spatial and frequency partitions maximising the average cell throughput while satisfying operator-defined system constraints such as using area-proportional frequency partitions or providing a certain fairness degree when allocating data-rates to the cell-centre and cell-edge UEs. In particular, three FFR-based designs are explored: 1) Fixed-spectrum-allocation-factor Design (FxD), 2) Area-proportional Design (ApD) and 3) Quality-constrained Design (QoSCD). In the first design, the use of a fixed frequency partition typically provides high spectral efficiencies at the cost of losing fairness when allocating data rates to cell-centre and cell-edge UEs. Using an area-proportional design allows a more equitable distribution of the spectrum that results in a higher degree of fairness between cell-centre and cell-edge UEs, at the cost of reducing the average cell throughput with respect to the FxD. Constraining the optimisation problem in order to ensure that the cell-edge

throughput is, at least, a fixed fraction of the cell-centre throughput, a certain fairness degree (i.e., a certain quality of service (QoS) to the cell-edge UEs) is enforced between cell-centre and cell-edge UEs while maximising the area spectral efficiency under these constraints.

The parameters used to pose the optimisation problems are the distance threshold ratio $\omega \triangleq R_{th}/R_m$ and the spectrum allocation factor $\rho \triangleq |\mathcal{F}_C|/|\mathcal{F}_T|$. It is expected that the appropriate selection of these parameters significantly affects the average cell throughput per subchannel [10] that, for convenience, is defined as³

$$\tau(\omega, \rho) \triangleq \frac{\eta(\omega, \rho)}{\Delta f |\mathcal{F}_T|} = \rho \tau^C(\omega, \rho) + \frac{1-\rho}{3} \tau^E(\omega, \rho), \quad (22)$$

measured in bps/Hz/subchannel, where $\tau^A(\omega, \rho) \triangleq \frac{\eta^A(\omega, \rho)}{\Delta f |\mathcal{F}_A|}$ is the average throughput per subchannel in cell region A .

A. Fixed-spectrum-allocation-factor Design

Under FxD, the spectrum allocation factor is fixed to $\rho = \rho_o$ (typically $\rho_o = 0.5$). Therefore, only the parameter ω remains to be optimised and the problem can be formulated as

$$\omega^* = \arg \max_{0 \leq \omega \leq 1} \rho_o \tau^C(\omega, \rho_o) + \frac{1-\rho_o}{3} \tau^E(\omega, \rho_o). \quad (23)$$

B. Area-proportional Design

In this case, the spectrum allocation factor is determined by the so-called area-proportional ratio as $\rho = \omega^2$ [10]. The optimisation problem can then be formulated as

$$\omega^* = \arg \max_{0 \leq \omega \leq 1} \omega^2 \tau^C(\omega, \omega^2) + \frac{1-\omega^2}{3} \tau^E(\omega, \omega^2), \quad (24)$$

where, obviously, $\rho^* = \omega^{*2}$.

C. QoS-constrained Design

In the QoSC approach, a QoS requirement q is stipulated enforcing that the guaranteed average throughput per-subchannel and per-edge UE is at least a fraction q of the average throughput per-subchannel and per-centre UE. Hence, the system parameters are adjusted to trade the data rates provided to the cell-centre UEs against those provided to the cell-edge UEs [10]. The corresponding optimisation problem can be formulated as

$$\begin{aligned} (\omega^*, \rho^*) = \arg \max_{0 \leq \omega, \rho \leq 1} & \rho \tau^C(\omega, \rho) + \frac{1-\rho}{3} \tau^E(\omega, \rho), \\ \text{subject to} & \tau_{mu}^E(\omega, \rho) \geq q \tau_{mu}^C(\omega, \rho), \end{aligned} \quad (25)$$

where the variables $\tau_{mu}^A(\omega, \rho)$ are defined in a different way for different scheduling policies. Assuming that the number of UEs served by the BS of interest is equal to $M_0 = k$ and as the PF and RR schedulers aim at offering either fair throughput values or fair access opportunities to the whole set of UEs, respectively, it seems reasonable to define the per-UE and per-subchannel average cell throughput as $\tau_{mu}^C(\omega, \rho) = \frac{\rho \tau^C(\omega, \rho)}{k P_C}$ for the cell-centre UEs and $\tau_{mu}^E(\omega, \rho) = \frac{(1-\rho) \tau^E(\omega, \rho)}{3k(1-P_C)}$ for the cell-edge UEs. When applying the MSINR scheduling strategy, however, as the scheduler aims at serving the UE experiencing the highest instantaneous SINR, it seems more

³Note that, in order to stress its dependency with respect to the optimisation parameters ρ and ω , the overall average throughput is represented as $\eta(\omega, \rho)$.

TABLE I: Network parameters

System parameter	Value
Cell radius	500 m
Minimum distance between BS and UEs	35 m
Antenna configuration	SISO
Transmit power of the BS	46 dBm
Antenna gain at the BS	14 dBi
Power spectral density of noise	-174 dBm/Hz
Receiver noise figure	7 dB
Total bandwidth	20 MHz
Total number of subcarriers	512
Path loss model	$15.3 + 37.6 \log_{10}(d)$ dB

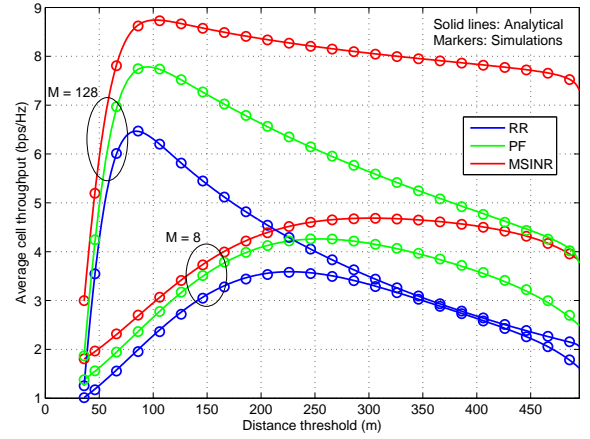


Fig. 2: Average cell throughput versus distance threshold R_{th} for different values of M (FxD).

reasonable to use the definitions $\tau_{mu}^C(\omega, \rho) = \rho \tau^C(\omega, \rho)$ and $\tau_{mu}^E(\omega, \rho) = \frac{(1-\rho) \tau^E(\omega, \rho)}{3}$ where note that the per-subchannel average throughput has not been averaged with respect to the average number of UEs in the corresponding cell region.

V. NUMERICAL RESULTS

In order to validate the proposed analytical framework while also providing valuable design guidelines, the 19-cell network shown in Fig. 1 is considered, where the cell of interest is surrounded by two rings of interfering BSs. As stated in previous sections, UEs are distributed over the coverage area using a PPP of normalised intensity λ_m (measured in UEs per area unit). For the sake of presentation clarity, although the analytical framework has been developed using the normalised intensity of the PPP, results in this section will be shown as a function of the average number of UEs per cell ($M \triangleq \pi \lambda_m R_m^2$). The main system parameters used to generate both the analytical and simulation results are based on [21] and summarised in Table I.

A. FxD-based FFR design

Figure 2 shows the average cell spectral efficiency (measured in bps/Hz) as a function of the distance threshold R_{th} for the FxD-based FFR design. Illustrating the system behaviour under different network deployments, analytical and simulation results are provided for the RR, MSINR and PF scheduling policies and with the average number of UEs per cell as parameter. The first point to highlight is the very good match between analytical and simulation

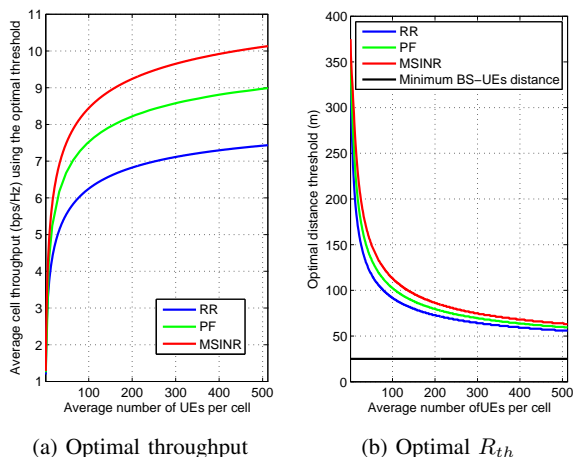


Fig. 3: Optimal average cell throughput and optimal distance threshold versus the average number of UEs per cell (FxD).

results, thus validating the analytical framework developed in Section III. Focusing now on performance issues, it can be observed that the average throughput increases with the average number of UEs per cell. This is basically due to two distinct effects. The first one, only exploited by the MSINR and PF scheduling rules, is caused by the larger degree of multiuser diversity provided by the increase of M , an effect that is more pronounced for the MSINR scheduler than for the PF scheduling rule because of the tradeoff that the PF scheduler establishes between the exploitation of multiuser diversity and the provision of fairness among UEs. The second effect, affecting all the schedulers but more noticeable when using the RR scheduler, is because increasing the average number of UEs per cell raises the probabilities of having at least one cell-centre UE and one cell-edge UE, hence reducing the probability of ending up with unassigned subcarriers and the consequent waste of resources.

We note that, having established the accuracy of the proposed analytical method, results shown in Figs. 3a and 3b depict only the theoretical outcomes, notwithstanding that they have also been duly replicated by means of simulations (not shown here for the sake of clarity). The optimal average cell throughput improvement caused by an increase in the average number of UEs per cell is shown in Fig. 3a. The corresponding optimal threshold radius R_{th} is shown in Fig. 3b. Note that, irrespective of the applied scheduling rule, the optimal distance threshold decreases when increasing the average number of UEs per cell, thus making the cell-edge region larger and causing the average cell-edge throughput to have a larger impact on the average cell throughput. Again this effect can be largely attributed to the fact that the more users in the system, the lower the probability of having unassigned subcarriers in the centre region. Finally, notice that, as in can be observed in Fig. 3b, the fairer the scheduling rule in use the smaller the optimal distance threshold becomes.

B. ApD-based FFR design

In ApD-based optimisation problems the spectrum allocation factor is selected as $\rho = \omega^2$. Using this setup, Figs. 4a, 4b and 4c show, for the MSINR, RR and PF scheduling

policies, respectively, the analytical and simulated average cell throughput as a function of the distance threshold ratio ω and with different configurations of the density of UEs per cell. As expected, due to making use of multiuser diversity, the PF and MSINR scheduling rules provide higher average throughput than the RR scheduler. This effect becomes even more evident as the density of UEs per cell increases.

As the MSINR scheduler does not consider the provision of fairness to UEs, it seems quite obvious that the optimal ApD will provide all the available resources to those UEs experiencing the higher SINR values, that is, the cell-centre UEs. This is the result that can be extracted from Fig. 4a where, irrespective of the network configuration, the optimal distance threshold ratio is $\omega^* = 1$, thus implying that the optimal ApD spectral allocation factor is $\rho^* = 1$. That is, when using MSINR in an ApD-based FFR-aided network, the optimal solution consists of using a full spectrum reuse strategy (cell-edge regions are eliminated and thus, all the subcarriers are allocated to the cell-centre area). When using the RR scheduling policy, and as it can be observed in Fig. 4b, there is an optimal distance threshold ratio around $\omega = 0.81$, that is rather insensitive to the density of UEs in the system. The optimal values of ω (and hence, those of ρ) are almost independent of M because the RR scheduler does not take advantage of the multiuser diversity. The performance of the PF scheduler, shown in Fig. 4c, resembles that of the MSINR in the sense that it tends to favour full spectrum reuse but notice that the attained average cell capacity is considerably lower than for the MSINR as some of the multiuser diversity is traded off by an increase in fairness among users.

C. QoS-based FFR design

To assess the performance of the QoS strategy under different QoS constraints, two different quality factors, $q = 0.02$ and $q = 0.2$, have been considered, corresponding to low and high throughput fairness requirements between cell-centre and cell-edge UEs. Analytical and simulated average throughput as a function of the distance threshold ratio ω and with different UE densities are shown in Fig. 5 for the three considered schedulers. Again, it is worth noting the good agreement between the simulated and analytical results. Also note that for these results, for each value of ω , the value of ρ maximising the average throughput has been used. The pairs (ω, ρ) leading to the maximum average throughput observed in these graphs are indeed the solutions to problem (25).

Similar to the other designs, the MSINR scheduler provides a clear average throughput performance advantage when compared to the PF and RR schedulers owing to the optimal exploitation of multiuser diversity. Logically, this performance advantage becomes more pronounced in environments with a high density of UEs. Increasing the QoS requirement q enforces a higher degree of fairness between cell-centre and cell-edge UEs at the cost of decreasing the spatially averaged throughput. In an ApD-based network using an MSINR scheduling rule we concluded that the full spectrum reuse scheme is the one providing the maximum average system throughput. The full spectrum reuse solution, that is $(\rho, \omega) = (1, 1)$, coincides with that obtained when designing an FFR-aided QoS-based network with $q = 0$. This is obviously a trivial non-realistic QoS constrained design, but it

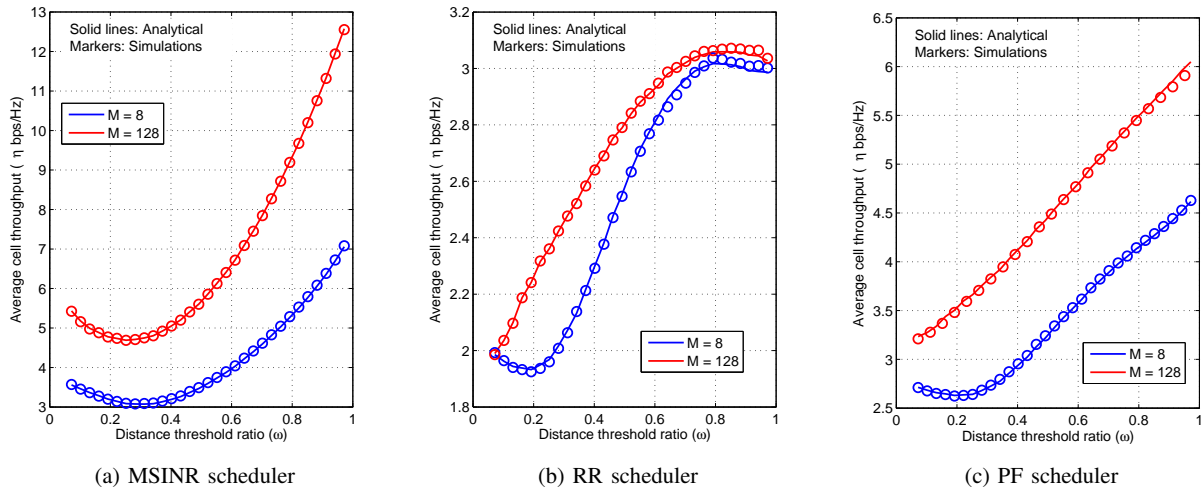


Fig. 4: Average cell throughput versus distance threshold ratio ω (assuming the use of $\rho = \omega^2$) for different values of the average number of UEs per cell M (ApD).

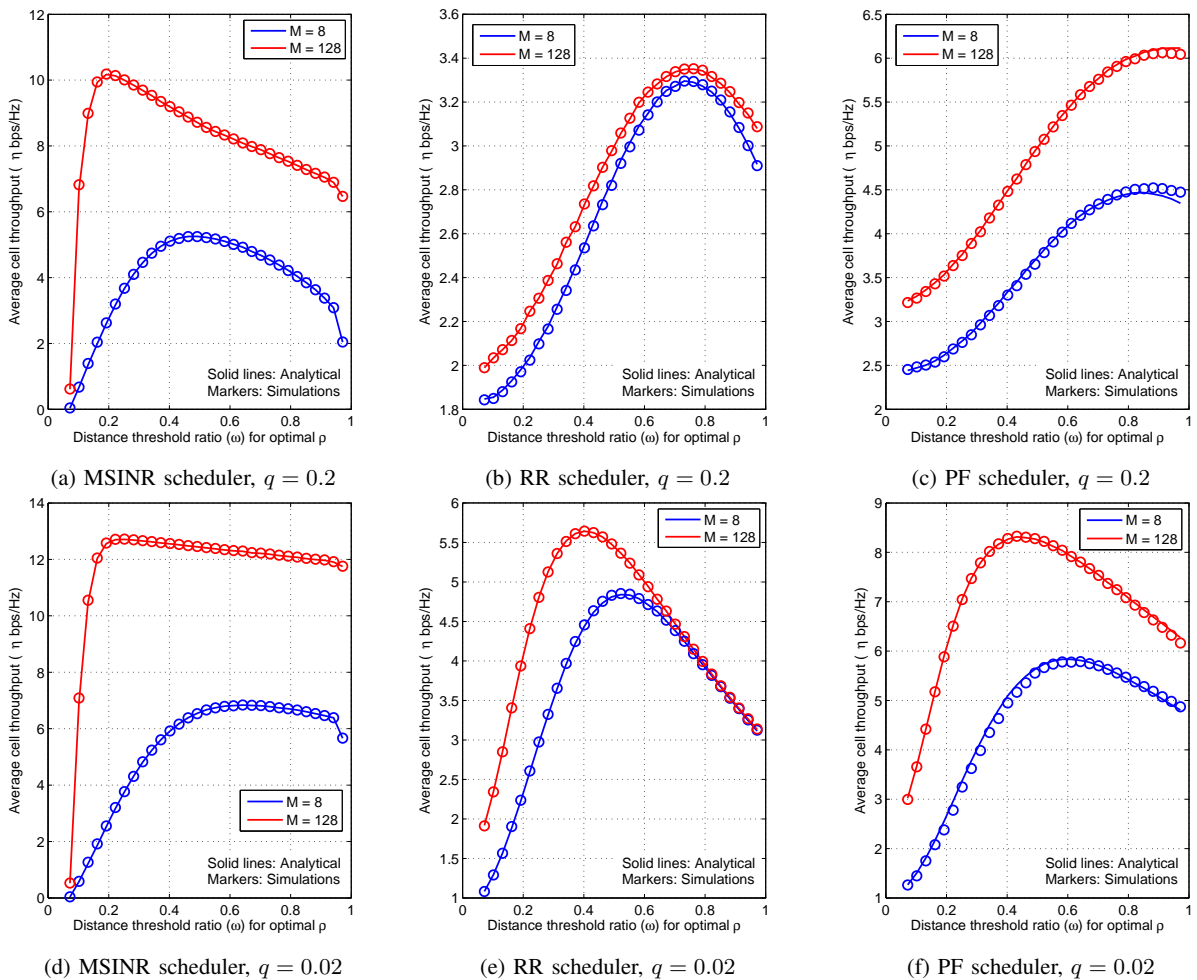


Fig. 5: Average cell throughput versus distance threshold ratio ω (assuming the use of the corresponding optimal value of ρ) for different values of the average number of UEs per cell M (QoSCD).

can help us understand the MSINR-related results presented in Figs. 5d and 5a. As the QoSCD strategy enforces a minimum degree of fairness between cell-centre and cell-edge UEs, the constrained design aiming at maximising the average system throughput may no longer match the full spectrum reuse scheme, however, the optimal spectrum allocation factor for $q = 0.02$ has been found to approach 95% of all the available spectral resources, that is, the optimal design tends to the full spectrum reuse strategy. Increasing the QoS requirement from $q = 0.02$ to $q = 0.2$, produces a notable decrease in both the optimal spectrum allocation factor and the optimal distance threshold ratio, thus increasing the amount of spatial and frequential resources allocated to the cell-edge UEs.

When using the RR scheduler, as it can be observed in Figs. 5b and 5e, the optimal distance threshold ratio increases with the quality factor q at the cost of a reduction in the average throughput. In fact, as the QoS constraint is defined as the requirement that $\tau_{mu}^E(\omega, \rho)$ exceeds a certain fraction of $\tau_{mu}^C(\omega, \rho)$, where remember that $\tau_{mu}^A(\omega, \rho)$ is defined as the average per-UE and per-subcarrier throughput in region A , strict QoS constraints in an RR-based network can only be achieved by increasing the cell-centre area (i.e., the number of cell-centre UEs) and the number of subcarriers allocated to the cell-centre UEs. As expected, the performance of the PF scheduler (shown in Figs. 5c and 5f) sits in between that of the MSINR and RR, where again it can be observed that increasing q leads to an increase in the optimum distance threshold, and as in the RR case, makes this optimum almost independent of the number of UEs in the system.

VI. CONCLUSION

This paper has introduced and validated an analytical framework to evaluate the performance of FFR-aided OFDMA networks. Interestingly, the proposed model is able to incorporate the effects of the scheduler and closed-form solutions have been derived for three popular scheduling rules, namely, MSINR, RR and PF. It has been shown that the average system throughput increases with the number of users as the probability of ending up with an empty FFR-defined area (i.e., cell-center or cell-edge) vanishes, thus avoiding the waste of radio resources. Remarkably, for the specific case of MSINR scheduling, and to a lesser extent for the PF rule, this throughput improvement is further accentuated by a greater exploitation of the multiuser diversity. Regarding the FFR design, it is worth stressing that the optimal threshold radius has been found to significantly depend on the number of users per cell and the scheduling rule implemented at the BS, thus highlighting the importance of analytical models, such as the one proposed in this paper, that allow a quick and accurate performance evaluation under the specific system conditions. Finally, different FFR designs have been considered that suitably dimension the central and edge FFR defined-regions and the amount of frequency resources assigned to each region. In particular, ApD and QoSCD designs have been shown to offer different operating points in the throughput vs fairness plane. Further work will concentrate on extending the analysis so as to encompass multi-tier OFDMA heterogeneous networks (e.g. macro and femto tiers) and the use of more sophisticated ICIC techniques (e.g., soft/adaptive frequency reuse, network MIMO).

ACKNOWLEDGEMENTS

Work supported by MINECO (Spanish Government) and FEDER under projects TEC2011-25446 and TEC2014-59255-C3-2-R, and the Conselleria d'Educació, Cultura i Universitats (Govern de les Illes Balears) under grant FPI/1538/2013 (co-financed by the European Social Fund).

REFERENCES

- [1] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*, 2nd ed. Elsevier Science, 2013.
- [2] A. S. Hamza, S. S. Khalifa, H. S. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1642–1670, 2013.
- [3] N. Saquib, E. Hossain, and D. I. Kim, "Fractional frequency reuse for interference management in LTE-Advanced HetNets," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 113–122, 2013.
- [4] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *IEEE International Conference on Communications (ICC)*, vol. 1, 1995, pp. 331–335.
- [5] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.
- [6] S. Shakkottai and A. Stolyar, *Scheduling algorithms for a mixture of real-time and non-real-time data in HDR*. Bell Laboratories, Lucent Technologies, 2000.
- [7] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4294–4305, December 2011.
- [8] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 996–1019, 2013.
- [9] R. W. Heath, M. Kountouris, and T. Bai, "Modeling heterogeneous network interference using Poisson point processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4114–4126, 2013.
- [10] F. Jin, R. Zhang, and L. Hanzo, "Fractional frequency reuse aided twin-layer femtocell networks: Analysis, design and optimization," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 2074–2085, 2013.
- [11] M. Assaad, "Optimal fractional frequency reuse (FFR) in multicellular OFDMA system," in *IEEE 68th Vehicular Technology Conference (VTC-Fall)*, 2008, pp. 1–5.
- [12] A. Najjar, N. Hamdi, and A. Bouallegue, "Efficient frequency reuse scheme for multi-cell OFDMA systems," in *IEEE Symposium on Computers and Communications (ISCC)*, 2009, pp. 261–265.
- [13] J. García-Morales, G. Femenias, and F. Riera-Palou, "Analytical performance evaluation of OFDMA-based heterogeneous cellular networks using FFR," in *IEEE 81st Vehicular Technology Conference (VTC-Spring)*, 2015.
- [14] G. Femenias and F. Riera-Palou, "Corrections to, and comments on, 'Throughput and Optimal Threshold for FFR Schemes in OFDMA Cellular Networks,'" *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2926 – 2928, 2015.
- [15] J.-G. Choi and S. Bahk, "Cell-throughput analysis of the proportional fair scheduler in the single-cell environment," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 766–778, 2007.
- [16] E. Liu and K. Leung, "Expected throughput of the proportional fair scheduling over Rayleigh fading channels," *IEEE Communications Letters*, vol. 14, no. 6, pp. 515–517, June 2010.
- [17] J. Wu, N. B. Mehta, A. F. Molisch, and J. Zhang, "Unified spectral efficiency analysis of cellular systems with channel-aware schedulers," *IEEE Transactions on Commun.*, vol. 59, no. 12, pp. 3463–3474, 2011.
- [18] D. Parruca, M. Grysla, S. Gortzen, and J. Gross, "Analytical model of proportional fair scheduling in interference-limited OFDMA/LTE networks," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sept 2013, pp. 1–7.
- [19] Z. Xu, G. Y. Li, C. Yang, and X. Zhu, "Throughput and optimal threshold for FFR schemes in OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2776–2785, 2012.
- [20] M. Maqbool, P. Godlewski, M. Coupechoux, and J.-M. Kéfil, "Analytical performance evaluation of various frequency reuse and scheduling schemes in cellular OFDMA networks," *Performance Evaluation*, vol. 67, no. 4, pp. 318–337, 2010.
- [21] G. TR36.921, "Home enode b (HeNB) radio frequency (RF) requirements analysis (release 9)," v9.0.0. Mar. 2010.

Propuesta de un algoritmo de encaminamiento para distribuir la exposición a los campos electromagnéticos en redes inalámbricas malladas

Julián Igareda[†], Luis Diez[†], Voichita Iancu[§], Emil Slusanschi[§], Ramón Agüero[†]

[†]Departamento de Ingeniería de Comunicación,
Universidad de Cantabria, 39005
julian.igareda@alumnos.unican.es,
{ldiez, ramon, luis}@tlmat.unican.es

[§]University Politehnica of Bucharest,
Splaiul Independentei nr.313
060042 București, Romania
{voichita.iancu, emil.slusanschi}@cs.pub.ro

Resumen—En este trabajo se presenta un nuevo modelado de redes multi-salto que tiene como objetivo el incluir métricas de exposición a campos electromagnéticos en el encaminamiento. Dada la necesidad de incluir diferentes métricas en las decisiones de creación de rutas, previo paso a la definición de nuevos protocolos de encaminamiento, se hace necesaria la elaboración de modelos de red y algoritmos adecuados que puedan ser posteriormente utilizados para medir el rendimiento de estos protocolos. En concreto, en este trabajo se propone un modelo de red adecuado para incluir en los costes de encaminamiento la exposición a los campos electromagnéticos; asimismo se ha identificado un algoritmo (*Cycle Canceling Algorithm*) capaz de solucionar el problema que emana del modelo. Además, se ha comparado el rendimiento de la solución propuesta frente a otros algoritmos en función de diferentes métricas.

Palabras Clave—Redes multi-salto, Campos Electromagnéticos, Cancelación de Ciclo

I. INTRODUCCIÓN

Resulta evidente que las comunicaciones inalámbricas han venido creciendo a un ritmo muy elevado en los últimos años, y se espera que siga siendo así en el medio plazo. A pesar de que no es razonable plantear un escenario diferente, han aparecido, asimismo, opiniones en las que se muestra cierta preocupación por el efecto negativo que la exposición de los usuarios a los campos electromagnéticos puede tener [1] [2]. Es por ello que resulta adecuado estudiar la posibilidad de mejorar los procedimientos actuales de gestión de las redes inalámbricas para que se pueda reducir dicha exposición sin que el usuario perciba un empeoramiento de la calidad de experiencia en sus servicios. Este es precisamente el *leitmotiv* del proyecto europeo LEXNET [3], que se centra en la propuesta de diferentes soluciones, en varios niveles, para lograr dicho objetivo.

Una de las alternativas de comunicación que mayor atención ha centrado recientemente son las topologías multi-salto. A pesar de que su origen data de 1997 (redes MANET), el escenario típico de aplicación ha variado sensiblemente, pasando de los escenarios bélicos y desastres naturales que originaron la investigación de este tipo de redes, a casos de uso que se centran en la expansión de la cobertura de estaciones base de manera económica y sencilla. Es importante destacar, por ejemplo, la relevancia de este tipo de despliegues en las cada vez más extendidas redes de sensores inalámbricos o el

papel que se les otorga en las nuevas especificaciones de la tecnología LTE (comunicaciones *Device-to-Device*).

Con el objetivo de reducir la exposición a campos electromagnéticos en este tipo de despliegues, se propuso en [4] el protocolo Reward Based Routing Protocol (RBRP), demostrándose que, utilizando un enfoque similar al empleado por estrategias de encaminamiento para balancear la carga, conseguía no solo distribuir de manera más ecuánime la exposición por todo el área en la que se desplegaba la red, sino que se producía además un descenso del valor medio. Sin embargo, el análisis se limitaba a un único flujo de información y todos los nodos utilizaban la misma potencia de transmisión. Además, la comparativa se realizaba con un protocolo para redes multi-salto tradicional, pero no era posible establecer si el resultado se acercaba o no al comportamiento óptimo.

Es por ello que en este trabajo se propone un algoritmo con el que poder corroborar el comportamiento del RBRP. Para ello se plantea un problema de optimización que busca encaminar la información (paquetes) por aquellas rutas en las que se incurra en una exposición menor, teniendo en cuenta, además de la que se generará por el mero hecho de estar utilizando un conjunto de nodos para reenviar paquetes, aquella que se corresponde a la que se ha ido acumulando anteriormente. Se añade, además, la posibilidad de modificar la potencia de transmisión de los nodos de manera dinámica, ya que se considera que es uno de los parámetros con mayor influencia en la exposición generada en este tipo de despliegues de red. Para ello se propone una transformación a realizar en el grafo que modela la red multi-salto. Posteriormente, para encontrar las rutas óptimas, se modifica el algoritmo Cycle Canceling Algorithm (CCA), adaptándolo para resolver el problema planteado. Los resultados obtenidos, al compararlos con un algoritmo de búsqueda de camino de coste mínimo evidencian que la solución propuesta consigue distribuir la exposición a campos electromagnéticos; se analiza asimismo el compromiso entre la reducción de la potencia de transmisión y la distribución de la exposición que se consigue.

El resto del documento se ha estructurado tal y como se recoge a continuación: en la Sección II se propone un modelo novedoso, basado en teoría de grafos, que se adapta a la problemática propuesta. A continuación, en la Sección III se plantea un problema de optimización que, basado en el

modelo de red anterior, optimiza conjuntamente la exposición acumulada en los nodos, reduciendo a la vez la potencia transmitida por los nodos; finalmente se propone un algoritmo para solucionarlo. La Sección IV describe los principales resultados obtenidos a partir de la evaluación del algoritmo, comparando su comportamiento con el de otras alternativas más tradicionales. Finalmente, en la Sección V se concluye el trabajo, resaltando las principales conclusiones del estudio, así como las líneas de investigación que se afrontarán a partir del mismo.

II. MODELADO DE LA RED

Se presenta en esta sección un modelo de red, basado en la teoría de grafos, que tiene en cuenta la exposición acumulada en los nodos y la potencia transmitida por estos. Su principal objetivo es la posterior definición de algoritmos para optimizar ambos parámetros de manera conjunta y que así sirvan como referencia para el análisis del comportamiento de protocolos de encaminamiento que consideren como métrica a optimizar la exposición a campos electromagnéticos generados por la transmisión de los paquetes.

La red se modela como un grafo $G = (\mathcal{V}, \mathcal{E})$, en el que \mathcal{V} son los vértices y \mathcal{E} representa el conjunto de arcos. Se consideran dos costes diferenciados en el grafo que, atendiendo al parámetro que representan, se imputan a los arcos o a los nodos que componen la red, tal y como se describe a continuación.

- Exposición acumulada K : este coste se asigna a cada nodo de la red, e indica el coste asociado a la exposición que dicho nodo ha generado hasta un instante concreto t . Su valor es independiente del tráfico que atraviesa en nodo el cada momento.
- Exposición por potencia P : está relacionado con la potencia necesaria para que un nodo transmita un paquete al siguiente salto, y se asocia, por tanto, a cada uno de los enlaces. Se considera que este coste crece de manera proporcional al número de flujos que atraviesan un arco del grafo en un momento determinado.

Debido a la diferente naturaleza de los parámetros definidos previamente, es necesario llevar a cabo una transformación del grafo inicial, de modo que se puedan utilizar modificaciones de algoritmos conocidos para poder encontrar los caminos que minimicen ambos costes de manera conjunta. En concreto, se introducirán arcos *virtuales* a los que atribuir el coste K de los nodos. Así, cada nodo se transforma en dos vértices virtuales, para poder discriminar entre el tráfico entrante y saliente. Dado un nodo V , éste se descompone en dos, V' y V'' , que son los puntos de entrada y salida del tráfico del nodo original V , respectivamente; éstos se conectan por un arco cuyo coste es igual al valor de la exposición acumulada en V hasta ese momento, K . De este modo, el nodo V' tiene los mismos arcos entrantes que tenía V y como único arco saliente el que lo conecta con V'' (con coste K). Del mismo modo, V'' tiene como arcos salientes los que tenía V , y como único entrante el que sale de V' .

La Figura 1 representa la transformación aplicada a un sencillo grafo de 3 nodos. Como puede observarse, los costes en los arcos se corresponden inicialmente con la potencia necesaria para comunicarse con los diferentes vecinos, escalado por el flujo que atraviesa el arco, véase la Figura 1a. Tal

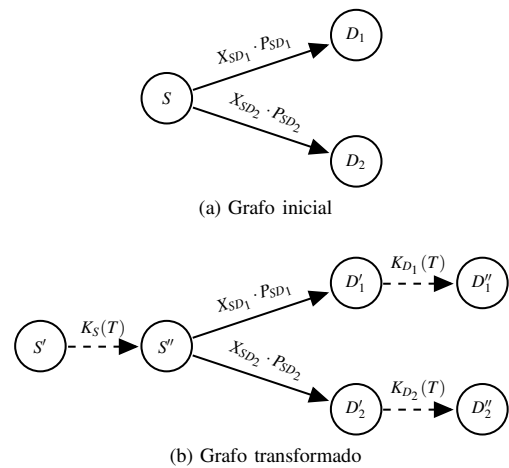


Figura 1: Ejemplo de transformación del grafo. X_{ij} y P_{ij} representan los valores de flujo de tráfico y potencia de transmisión entre los nodos i y j . $K_i(T)$ representa el coste relativo a la exposición acumulada en el nodo i en el momento T .

como muestra la Figura 1b, tras aplicar la transformación propuesta, aparecen nuevos arcos, con el coste de la exposición acumulada en los nodos originales. Por otro lado, se puede observar que el tamaño del grafo inicial aumenta, tanto en el número de vértices, como de arcos. Si se parte de un grafo $G = (\mathcal{V}, \mathcal{E})$, se obtiene un grafo transformado $G' = (\mathcal{V}', \mathcal{E}')$ en el que $|\mathcal{V}'| = 2 \cdot |\mathcal{V}|$ y $|\mathcal{E}'| = |\mathcal{V}| + |\mathcal{E}|$.

III. FORMULACIÓN DEL PROBLEMA

Una vez descrita la transformación que se lleva a cabo sobre el grafo, esta sección presenta el problema de optimización que se plantea para establecer las rutas óptimas, y que tiene en cuenta tanto la exposición acumulada como la potencia de transmisión de los nodos. Asimismo, se analizarán un conjunto de algoritmos existentes, de manera que se pueda identificar aquel que se tomará como base para resolver el problema, y para establecer una referencia a la hora de establecer el protocolo de encaminamiento.

Dada la estructura del grafo de partida, en el que el coste relacionado con la potencia de transmisión es proporcional al flujo total en cada enlace, el problema se plantea inicialmente como un Minimum Cost Flow (MCF), que tiene como finalidad el distribuir un flujo entre un conjunto de nodos fuente y destino, minimizando el coste global.

De forma analítica, el problema de puede representar como se muestra a continuación.

$$\begin{aligned} \text{mín} \quad & \sum_{\forall(i,j) \in \mathcal{E}} c_{ij} \cdot x_{ij} & (1) \\ \text{s.t.} \quad & \forall i \in \mathcal{N} \quad \sum_{j:(i,j) \in \mathcal{E}} x_{ij} - \sum_{j:(j,i) \in \mathcal{E}} x_{ji} = b(i) & (2a) \\ & \forall(i,j) \in \mathcal{E} \quad u_{ij} \geq x_{ij} \geq 0 & (2b) \\ & \sum_{i=1}^n b(i) = 0 & (2c) \end{aligned}$$

Donde c_{ij} y u_{ij} representan el coste por unidad de flujo y la capacidad del enlace $(i,j) \in \mathcal{E}$, respectivamente. Por otro

lado, $b(i)$ se corresponde con el flujo originado ($b(i) > 0$) o recibido ($b(i) < 0$) en cada nodo. Finalmente, x_{ij} (variable a optimizar) representa el flujo que se envía por cada enlace $(i, j) \in \mathcal{E}$. Así, la restricción 2a asegura que el balance de flujo en cada nodo coincida con la generación/demanda del mismo (cuando es origen o destino), siendo nulo si se trata de un nodo intermedio. La restricción 2b modela que el flujo en cada enlace no supere la capacidad del mismo; finalmente, para asegurar la solución del problema, se establece que el balance total de flujo en la red sea nulo, a través de la restricción 2c.

Existen diferentes soluciones a este problema, atendiendo a las propiedades que se pretendan potenciar. En este trabajo, dado que se centra en el modelado y no tanto en el algoritmo para resolverlo, se ha optado por una solución tradicional como es el Cycle Canceling Algorithm. En este caso se hace uso de la propiedad que establece que una solución x^* es óptima si, y solo si, el grafo residual $G(x^*)$ no contiene ningún ciclo negativo [5]. La funcionalidad básica del algoritmo se muestra en 1, en la que r_{ij} representa la capacidad residual del arco (i, j) . El algoritmo parte de un grafo inicial $G(\mathcal{N}, \mathcal{E})$, donde \mathcal{N} indica el conjunto de nodos y \mathcal{E} el de arcos, en el que la distribución inicial de flujo se realiza de manera aleatoria. En cada iteración el algoritmo encuentra un ciclo negativo W y lo cancela aumentando la cantidad de flujo en ese ciclo. Finalmente la solución óptima se corresponde a la distribución de flujo cuando no se encuentran más flujos negativos. Una consideración importante para la su resolución es que este algoritmo se aplica tradicionalmente sobre redes con una única fuente de tráfico (con balance de flujo negativo) y un destino (con balance de flujo positivo). Para poder considerar múltiples fuentes y destinos se definió un par origen/destino virtuales, de modo que la capacidad de los enlaces desde el origen virtual a cada una de las fuentes sea 1 (esto es, solo puede utilizarse por un flujo), asegurando así que cada fuente original de tráfico genere efectivamente una unidad de flujo. En el caso del destino virtual, no se ha limitado la capacidad de los enlaces que le unen con los nodos destinos originales, dado que no se ha impuesto ninguna restricción para lograr cierto balanceo de carga en los destinos.

Algoritmo 1 Algoritmo *Cycle Canceling*

Entrada: $G(\mathcal{N}, \mathcal{E})$

- 1: Establecer un flujo x posible en la red
 - 2: **MIENTRAS** $G(x)$ contenga un ciclo negativo **hacer**
 - 3: Identificar un ciclo negativo W
 - 4: $\delta := \min\{r_{ij} : (i, j) \in W\}$
 - 5: Aumentar δ unidades en flujo en el ciclo W y actualizar $G(x)$
 - 6: **FIN MIENTRAS**
-

Cabe destacar que, debido a la transformación del grafo presentada en la Sección II, aparecen enlaces cuyos costes no crecen linealmente con el flujo que atraviesa dicho arco, sino que permanecen constantes, ya que modelan la exposición acumulada hasta ese momento, por lo que la función de utilidad quedaría como recoge la Ecuación 3.

$$\min \left(\sum_{\forall(i,j) \in E} c_{ij} \cdot x_{ij} + \sum_{\forall(i,j) \in E} k_{ij} \right) \quad (3)$$

Esta modificación se ha considerado en la implementación del algoritmo, por lo que la eficacia de la solución obtenida tiene cierto carácter empírico. En la Sección IV se observará que, efectivamente, el algoritmo implementado se comporta de manera adecuada. En términos de implementación, se ha modulado el valor de la exposición acumulada K , de manera que, al multiplicarlo por el flujo que atraviesa el arco, siempre presente el mismo coste.

IV. ANÁLISIS DE LOS RESULTADOS

En esta sección se presentan los principales resultados obtenidos durante la evaluación del algoritmo propuesto. El análisis que se expone se estructura en dos partes diferenciadas: la Sección IV-A describe el comportamiento puesto de manifiesto por el algoritmo sobre topologías de red estáticas, lo que permitirá comparar sus prestaciones con alternativas que se basen en métricas tradicionales, en concreto con un algoritmo cuyo objetivo es la minimización de la potencia total transmitida, utilizando para ello una solución clásica de coste mínimo, en el que el coste de cada salto se corresponda con la potencia necesaria para lograr transmitir los paquetes al siguiente salto. Por otro lado, la Sección IV-B estudia el comportamiento de la solución propuesta en escenarios más complejos, en los que la evolución temporal de la exposición en los nodos adquiere una mayor relevancia; también se analizará la influencia de las posibles configuraciones del algoritmo desarrollado sobre el comportamiento que ofrece.

En ambos casos, el procedimiento consiste en ir aplicando los diferentes algoritmos sobre una red de manera secuencial; para ello se ‘emula’ la generación de flujos de tráfico (asumiendo que todos ellos son idénticos) y, por tanto, cada ejecución se corresponde con la búsqueda de rutas en instantes discretos, actualizando la situación concreta de la red (exposición acumulada) antes de utilizar los algoritmos, reflejando la evolución temporal que se seguiría por un protocolo. Esta metodología permite utilizar una implementación *a medida*, que considere la evolución temporal debida a los diferentes flujos de tráfico, sin la necesidad de utilizar herramientas más pesadas, que ralentizarían de manera notable el tiempo necesario para llevar a cabo el análisis.

Hay que tener en cuenta que la exposición acumulada se corresponde con un parámetro físico (típicamente medido en valores de campo eléctrico), mientras que en el algoritmo propuesto se modela a través del parámetro K , que bien podría estimarse a partir de los paquetes transmitidos por un nodo durante una ventana temporal concreta. Considerando que los flujos que se están utilizando son iguales, se ha establecido que, en cada ejecución, todos los nodos que fueron seleccionados para formar parte de una ruta en la ejecución anterior (a excepción del destino) incrementen su exposición en un valor constante de E , antes de proceder a ejecutar el algoritmo en la iteración actual.

IV-A. Configuración estática

En este primer análisis se evalúa si el algoritmo propuesto es capaz de encontrar rutas que optimicen de manera conjunta las métricas de potencia transmitida y de exposición acumulada.

En concreto, se han estudiado escenarios en los que se han desplegado aleatoriamente (según un proceso de *Poisson*) 50

Tabla I: Configuración topología estática

	Topología			
	α	β	γ	δ
Fuentes	{2,4}	{2,4}	{2,4,6,8}	{2,4,6,8}
Área m^2	120×120	130×130	140×140	150×150

nodos en superficies cuadradas, cuyo lado se ha ido variando paulatinamente, para modificar la densidad de nodos y, por tanto, el grado de la conectividad del grafo subyacente; se ha mantenido constante el radio de cobertura de los nodos en 15 m. La Tabla I resume los parámetros de este primer análisis; en cada topología (correspondiéndose con cada valor para el lado del área) se han analizado los resultados que se observaron al modificar el número de fuentes (y, por tanto, de flujos). En todos los escenarios se han fijado 4 posibles destinos, que se corresponden con elementos de acceso a la red, de manera que sea indiferente que la ruta se conecte con uno u otro destino, siempre que se consiga establecer un camino hacia cualquiera de ellos. Esta suposición, que se alinea de manera adecuada con la definición del destino virtual descrita anteriormente, podría reflejar una situación real en la que un usuario simplemente pretenda encontrar una ruta válida hacia algún *gateway* disponible, sin que tenga ninguna preferencia a priori por alguno de ellos.

Para analizar la evolución de la exposición se ha ejecutado el algoritmo, en todos los casos, 100 veces consecutivas, modificando el valor de exposición en función de las rutas establecidas en la ejecución anterior; así, los nodos que hayan formado parte de una ruta durante T unidades temporales durante el experimento presentarán una exposición acumulada $E \cdot T$. Además, para asegurar la validez estadística de los resultados, se han llevado a cabo 100 experimentos independientes para cada una de las configuraciones. Se ha asumido que el coste asociado a la potencia de transmisión es proporcional a ésta y que la potencia que transmiten los nodos es, a su vez, proporcional a la distancia entre ellos, siendo su valor máximo 16 dBm. Por tanto, el cálculo del coste de la potencia en el enlace (i, j) se puede expresar como $P_{ij} = \frac{d_{ij}}{15} \cdot 16$, donde d_{ij} se corresponde con la distancia entre los nodos. Por su parte la exposición provocada por cada flujo que atraviesa un nodo durante una ventana temporal incrementa el coste K de dicho nodo en 5 unidades.

En primer lugar se ha estudiado la exposición acumulada en los diferentes nodos a lo largo de un experimento, comparando el comportamiento del algoritmo propuesto con el que establece las rutas de mínima potencia. La Figura 2 muestra la función de probabilidad de la exposición de los nodos para ambos algoritmos, utilizando la topología α con 2 y 4 fuentes. El valor $E \cdot T$ se corresponde con la exposición que se acumularía en un nodo que siempre hubiera formado parte de una ruta durante el experimento de duración T pasos, mientras que sus múltiplos de este valor reflejan situaciones en las que un nodo ha formado parte de más de una ruta. Como puede observarse, el algoritmo de mínima potencia, con un carácter claramente estático únicamente presenta una probabilidad mayor de 0 en los múltiplos de $E \cdot T$. Por su parte, el algoritmo propuesto consigue balancear la exposición (y de manera colateral la carga); la consecuencia más visible en la

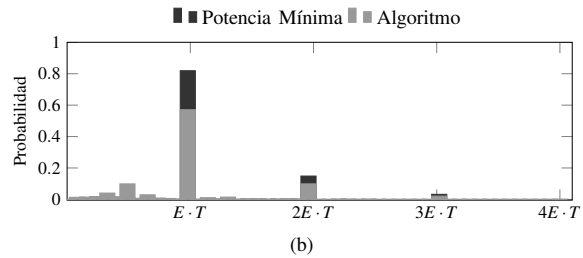
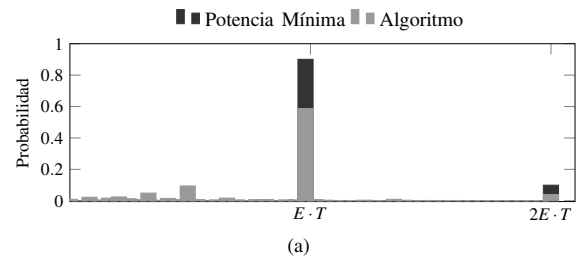


Figura 2: Función de probabilidad de la exposición en cada nodo en la topología α , La Figura 2a corresponde a la configuración con 2 fuentes y la Figura 2b a la de 4 fuentes

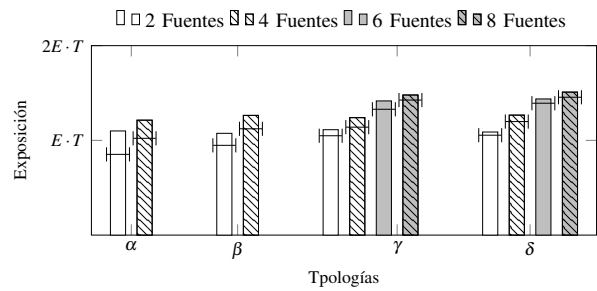


Figura 3: Exposición media por experimento. Los marcadores indican el valor que toma el algoritmo en cada configuración, las columnas indican el valor del algoritmo de mínima potencia

reducción que se produce en la probabilidad de que un nodo participe, de manera continuada, en una única ruta. Cuando se utilizan 2 fuentes de tráfico, la reducción es cercana al 30%, mientras que al aumentar el número de fuentes a 4 tiene como consecuencia que haya más nodos generando exposición y, por tanto, la ganancia es ligeramente inferior, situándose cerca del 20%.

Una vez estudiado el efecto que el algoritmo tiene sobre la exposición, se ha extendido el análisis, considerando varias topologías y diferentes cargas de tráfico (número de fuentes/flujos). La Figura 3 muestra el valor medio de la exposición (global, considerando todos los nodos) tras llevar a cabo 100 experimentos independientes. De acuerdo a los resultados observados, se puede decir que el beneficio del algoritmo, respecto a uno que establezca los caminos de mínima potencia se atenúa al disminuir la densidad de nodos. La figura muestra que con la ganancia con la topología menos densa (δ) es notablemente inferior a la observada sobre el escenario más denso (α). Sin embargo, la influencia del número de fuentes/flujos no es tan relevante, viéndose que prácticamente no tiene influencia sobre la exposición.

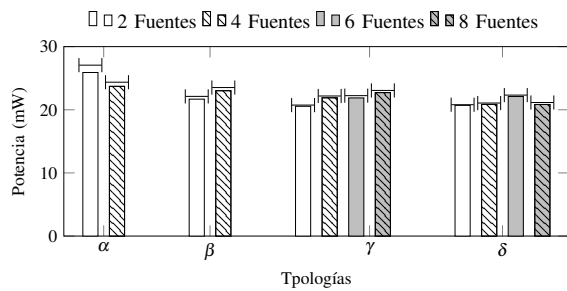


Figura 4: Potencia media por ruta. Los marcadores indican el valor que toma el algoritmo en cada configuración, las columnas indican el valor del algoritmo de mínima potencia

Teniendo en cuenta que la exposición generada por el conjunto de toda la red está directamente relacionada con lo potencia total transmitida, la caracterización del algoritmo también incluye una comparación de la potencia transmitida respecto a la óptima, que se ha obtenido utilizando un algoritmo tradicional búsqueda de coste (potencia) mínimo. Para ello, y de manera ilustrativa, se ha utilizado una relación lineal de la potencia necesaria por enlace con la distancia entre sus dos nodos extremos, con un constante de proporcionalidad de $1\text{ mW}/m$, de manera que la potencia máxima que se podría tener por enlace es la correspondiente a la distancia máxima (cobertura), esto es 15 mW .

Para estudiar el comportamiento de los dos algoritmos en relación a la potencia transmitida, la Figura 4 representa el valor medio de la potencia de transmisión total para todas las rutas que se establecen por escenario. Hay que tener en cuenta que, mientras el algoritmo de mínima potencia mantiene las mismas rutas durante todo el experimento, la solución propuesta en este trabajo sí que podría hacerlo, de manera que se tiene que tener en cuenta el tiempo que una ruta permanece activa para acometer esta caracterización. En general, se puede observar que como consecuencia de la distribución de la exposición, se produce una pequeña penalización en términos de potencia, que se vuelve menos apreciable a medida que la densidad de los escenarios disminuye. Al considerar conjuntamente los dos últimos resultados se puede concluir que, para este tipo de escenarios estáticos, cuanto mayor sea la ganancia en términos de exposición acumulada, mayor será la penalización observada en la potencia de transmisión necesaria.

IV-B. Configuración dinámica

En esta sección se estudiará el comportamiento del algoritmo sobre escenarios más complejos, en los que se va modificando dinámicamente el conjunto de nodos fuente durante un mismo experimento. En concreto, se ha hecho uso de una superficie con dimensiones $140 \times 140\text{ m}^2$ sobre la que se han situado 9 posibles destinos (*gateways*) en un *array*, de tal forma que cubran el mayor área posible. Además, se han desplegado 50 nodos entre los que se seleccionan 4 fuentes de forma aleatoria. Al igual que en la caracterización anterior, se ha fijado el radio de cobertura de todos los nodos a 15 m .

Se han realizado experimentos con una duración de $T = 100$ pasos temporales, modificando los nodos que actúan como fuente cada 25 pasos temporales. Nuevamente, la exposición

Tabla II: Configuración de costes

Parámetro de coste	Valor
P	distancia (m)
Δk	{3, 6, 9, 12, 15}
∇k	{0, ..., Δk }

acumulada en cada nodo se incrementa en E unidades por cada flujo de tráfico que lo atraviesa. A diferencia del estudio anterior, en que el valor del coste K se mantenía constante respecto al incremento de la exposición, en este caso se analizará cómo afecta al comportamiento del algoritmo y, en concreto, cuál debería ser la relación entre los costes asociados a la potencia de transmisión y a la exposición acumulada. Los resultados obtenidos podrán ser utilizados posteriormente para calibrar la configuración de ambos parámetros en un protocolo de encaminamiento real. Así, si se supusiera que la exposición acumulada es siempre creciente, se alcanzaría una situación en el que el criterio de minimización de la potencia transmitida sería despreciable frente al de la exposición acumulada, por lo que no se tendría en cuenta a la hora de encontrar las rutas óptimas.

La influencia de la relación entre los costes K (exposición) y P (potencia) se ha analizado con dos enfoque complementarios: en primer lugar se ha modificado únicamente el peso relativo a la exposición acumulada, manteniendo constante el que afecta a la potencia; en segundo lugar se ha estudiado el comportamiento del algoritmo cuando se aplica una función de *envejecimiento* a la exposición acumulada en los nodos, de manera que el coste correspondiente se reduzca cuando los nodos no estén participando de manera activa en ninguna ruta.

Las diferentes configuraciones que se han empleado para los costes se resumen en la Tabla II, en la que P se corresponde al coste relativo a la potencia, que se calcula igual que en el análisis anterior. Respecto al parámetro K , este se calcula como el valor anterior más un incremento Δk por cada una de las rutas en las que el nodo en concreto participa; si un nodo no ha estado activo en ninguna ruta en la última ventana temporal, el coste se decremента una cantidad ∇k . Como se puede apreciar en la Tabla II, los valores utilizados para ∇ siempre son inferiores a los de Δ . De nuevo se han realizado 100 experimentos independientes para cada configuración, asegurando así la validez estadística de los resultados.

En primer lugar se ha estudiado la exposición siguiendo un procedimiento análogo al de la Sección IV-A. La Figura 5 muestra el valor medio de la exposición promedio en cada experimento, considerando únicamente aquellos nodos que generan exposición (los que participan en las rutas que se establecen). En este caso, si un nodo formara parte de las 4 rutas activas a lo largo de todo el experimento, su exposición final tendría un valor de $4E \cdot T$.

Como puede observarse en la Figura 5, al variar los orígenes de los flujos de datos a lo largo del experimento se reduce la ganancia en términos de exposición, comparada con los escenarios estáticos que se estudiaron anteriormente. Sin embargo, se observa que la función de *envejecimiento* que se ha incorporado tiene un efecto positivo, ya que se consigue así que las rutas con menor número de saltos vuelvan a ser seleccionadas más rápidamente, lo que permite reducir

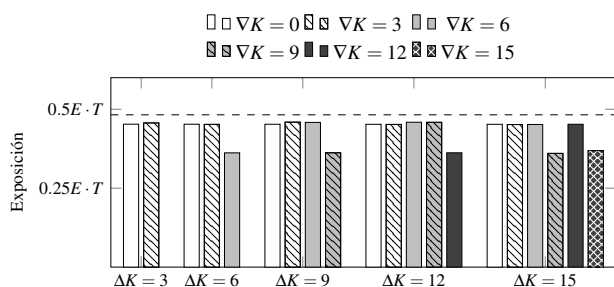


Figura 5: Exposición media al final de un experimento en cada configuración. La línea punteada marca el valor con el algoritmo de mínima potencia

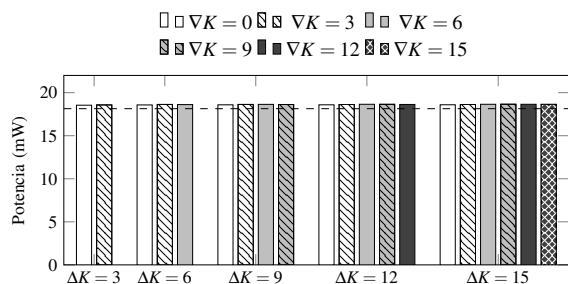


Figura 6: Potencia

la exposición. Además, los resultados ponen de manifiesto que sería recomendable que el coste asociado a la exposición tuviera un valor elevado en relación al correspondiente a la potencia de transmisión; en concreto, si este último se fija en 15 unidades, el correspondiente al crecimiento de exposición debería ser superior a 6.

Por otro lado, se ha vuelto a analizar la influencia del algoritmo sobre la potencia transmitida por ruta. La Figura 6 ilustra el valor medio de la potencia de transmisión por ruta por unidad de tiempo, comparando los resultados observados al utilizar el algoritmo que garantiza la mínima potencia con cada una de las configuraciones del algoritmo propuesto en este trabajo. Se observa que el hecho de que los escenarios tengan un carácter más dinámico tiene como consecuencia que todas las configuraciones se comporten de forma muy similar y que, además, los resultados de la solución propuesta en este trabajo (que busca, además, mejorar la distribución de la exposición) son muy similares a los que consigue el algoritmo que busca minimizar la potencia transmitida.

V. CONCLUSIÓN

Una de las consecuencias de la cada vez mayor presencia de las redes de comunicaciones móviles es la creciente preocupación de los usuarios por la exposición a los campos electromagnéticos que generan. Este es un aspecto al que todavía no se le ha dado mucha importancia, pero parece razonable pensar que sería interesante incluir este parámetro en el diseño de algoritmos, técnicas y protocolos a utilizar sobre redes inalámbricas. Este es uno de los principales objetivos del proyecto europeo LEXNET y en este trabajo se ha presentado una de las líneas de investigación que se han seguido en el mismo, en concreto para mejorar el comportamiento sobre redes inalámbricas multi-salto.

Se ha presentado un algoritmo que permite optimizar tanto la potencia de transmisión necesaria en cada ruta como la exposición acumulada por nodo. Para ello se ha planteado una transformación del grafo original, incorporando una serie de nodos virtuales, que permiten tener en cuenta el criterio que minimiza la exposición acumulada. El algoritmo propuesto se basa inicialmente en el CCA y los análisis que se han llevado a cabo, comparándole con una solución tradicional de búsqueda de caminos de coste mínimo, ponen de manifiesto que logra conseguir el objetivo para el que fue diseñado.

Los siguientes pasos a dar en esta línea de investigación son principalmente dos. En primer lugar se extenderá el análisis del protocolo RBRP, con la posibilidad de modificar la potencia de transmisión de manera dinámica y se contrastarán su rendimiento con los resultados aportados por el algoritmo propuesto en este trabajo. Por otro lado se elaborará un marco teórico que permita analizar de manera más amplia la métrica que se pretende mejorar, para lo que se planteará una formulación matemática formal de la exposición; con ella se podrá cuantificar, de manera más precisa, el comportamiento de las diferentes soluciones (algoritmos/protocolos) propuestas.

AGRADECIMIENTOS

Los autores agradecen la financiación de la Comisión Europea, en el marco del proyecto LEXNET, perteneciente al 7º Programa Marco (GA nº 318273). Asimismo, Ramón Agüero querría expresar su agradecimiento al Gobierno de España por su financiación en el proyecto “Connectivity as a Service: Access for the Internet of the Future”, COSAIF (TEC2012-38574-C02-02).

REFERENCIAS

- [1] M. Tesanovic, E. Conil, A. De Domenico, R. Aguero, F. Freudenstein, L. Correia, S. Bories, L. Martens, P. Wiedemann, and J. Wiart, “The LEXNET Project: Wireless Networks and EMF: Paving the Way for Low-EMF Networks of the Future,” *Vehicular Technology Magazine, IEEE*, vol. 9, no. 2, pp. 20–28, June 2014.
- [2] TNS Opinion & Social. (2010, June) Eurobarometer 73.3, Electromagnetic fields. [Online]. Available: http://ec.europa.eu/public_opinion/archives/ebs/ebs_347_en.pdf
- [3] (2012) Low electromagnetic field exposure networks. [Online]. Available: <http://www.lexnet-project.eu/>
- [4] V. Iancu, L. Diez, L. Rodriguez de Lope, E. Slusanschi, and R. Aguero, “A reward-based routing protocol to reduce the emf exposure over wireless mesh networks,” in *Wireless Days (WD), 2014 IFIP*, Nov 2014, pp. 1–4.
- [5] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

Análisis del Rendimiento de Técnicas de Network Coding y Encaminamiento Oportunista Sobre Redes Malladas Inalámbricas

Pablo Garrido, David Gómez, Ramón Agüero y Luis Muñoz

Universidad de Cantabria, Santander, España
{pgarrido, dgomez, ramon, luis}@tlmat.unican.es

Resumen—Este trabajo plantea como punto de partida el mal comportamiento que el protocolo *TCP* muestra cuando se utiliza sobre canales inalámbricos, medio de transmisión fundamental en las comunicaciones del siglo XXI, especialmente cuando sus condiciones son malas. Como solución, se propone una alternativa novedosa, basada en la combinación de un protocolo *Network Coding (NC) intra-flujo* con diferentes soluciones de encaminamiento oportunista, las cuales, gracias a la naturaleza *broadcast* del medio radio, permiten beneficiarse de la aparición de múltiples caminos entre un origen y un destino. Tras una extensa campaña de simulación sobre la plataforma *ns-3*, los resultados muestran una clara mejora con respecto al rendimiento observado al utilizar la solución tradicional (protocolo *TCP*), tanto en escenarios canónicos como en despliegues aleatorios.

Palabras Clave—Redes malladas inalámbricas, *TCP*, *Network Coding*, Codificación Lineal Aleatoria, Encaminamiento oportunista

I. INTRODUCCIÓN

Las redes inalámbricas se han convertido en la alternativa más extendida a la hora de proporcionar acceso a la red dorsal (conexiones de la última milla), gracias a la mejora continua de las tecnologías subyacentes, entre las que destacan las redes celulares y las redes de área local inalámbricas o *Wireless Local Area Networks (WLANs)*. Además, merece la pena destacar la presencia, cada vez más relevante, de las redes malladas inalámbricas, en las que los nodos pueden formar topologías multi-salto para extender la cobertura en determinadas zonas, de manera rápida y económica. Entre sus principales aplicaciones, destacan las redes de sensores o las comunicaciones *device-to-device*, definidas en el marco de la *5G*.

Sin embargo, la evolución que se ha visto en los niveles inferiores no ha sido correspondida en la capa de transporte, en la que *TCP* se mantiene, desde los orígenes de Internet, como la solución de facto. Este protocolo fue diseñado para trabajar principalmente sobre redes cableadas, y su comportamiento no es apropiado ante la presencia de errores aleatorios, que aparecen durante una transmisión sobre un canal inalámbrico real. De hecho, la reacción de sus mecanismos de control de la flujo será la misma que la producida tras eventos de congestión de red, provocando una grave infrautilización de los recursos disponibles.

Son numerosos los trabajos que han buscado en las últimas décadas solucionar estas deficiencias, ya sea a través de modificaciones directas en los mecanismos de control de flujo y congestión del propio protocolo *TCP* o, por otro lado,

planteando enfoques basados en diferentes alternativas *cross-layer*, que buscan reducir la influencia de los medios físicos hostiles sobre el rendimiento. Este trabajo se centra en la combinación de dos de las soluciones más prometedoras en el panorama actual; así, se parte de un esquema de codificación *Network Coding (NC) intra-flujo* sobre el que se sitúa una solución de encaminamiento oportunista, de manera que se aproveche la diversidad espacial que ofrece el canal inalámbrico. Como resultado, el servicio ofrecido es equivalente al que proporciona una comunicación orientada a la conexión, como *TCP*, ya que se garantiza la recepción ordenada de la información en la aplicación de los nodos destino. Para evaluar el comportamiento de la solución propuesta, se ha llevado a cabo una profunda campaña de simulación sobre la plataforma *ns-3*.

El documento se ha estructurado tal y como se describe a continuación: la Sección II describe algunas de las contribuciones más relevantes en las técnicas que se plantean en la solución propuesta, que es descrita en la Sección III, detallando las diferentes posibilidades de configuración de la misma. La Sección IV presenta los escenarios utilizados durante la campaña de simulación realizada y describe los principales resultados obtenidos. Finalmente, la Sección V concluye el documento, resumiendo las principales líneas de investigación que se potenciarán en el futuro.

II. TRABAJOS RELACIONADOS

TCP fue diseñado asumiendo que los errores eran consecuencia de la congestión de los nodos intermedios, y no de las interferencias o un canal de baja calidad. Así, los mecanismos empleados por *TCP* cuando malinterpreta la causa de una pérdida, generan una gran pérdida del rendimiento; en concreto, en redes malladas inalámbricas [1], éste se ve gravemente perjudicado. Muchos trabajos han estudiado el impacto de los canales inalámbricos sobre el rendimiento de *TCP* [2], [3] y también se han llevado a cabo propuestas para solventar las limitaciones observadas; entre ellas *Network Coding (NC)* aparece como una de las más prometedoras.

Las propuestas que han surgido se pueden dividir en dos grandes categorías. Por un lado, las *Inter-flujo*, en las que los nodos intermedios combinan paquetes pertenecientes a diferentes flujos de información. Dentro de este grupo destaca *COPE* [4], una arquitectura que propone la utilización de una operación *XOR* para mezclar los paquetes, y que presenta una

importante mejoría en el rendimiento sobre topologías canónicas. Sin embargo, otros trabajos han analizado el rendimiento de estas soluciones en condiciones no idóneas, mostrando que la ganancia es muy limitada [5], [6].

Otra alternativa son las técnicas que se conocen en la literatura como *Intra-flujo*; en ellas se combinan paquetes pertenecientes al mismo flujo. Cabe destacar el trabajo de Chachulski *et al.* [7] donde se propone *MAC independent Opportunistic Routing (MORE)*, un protocolo que combina técnicas de encaminamiento oportunista [8] y *NC*. El trabajo que se presenta aquí se basa en el protocolo *MORE*, pero lo combina con el protocolo *UDP*, y realiza un estudio sistemático de su rendimiento sobre redes complejas (despliegues aleatorios).

Otra propuesta que mejora el rendimiento de *TCP* sobre redes malladas inalámbricas es el protocolo *TCP/NC*, descrito por Sundararajan *et al.* en [9]; se añade una nueva capa entre *TCP* y *IP*, que se aprovecha de las ventajas de los mecanismos de *TCP*, que aseguran conexiones seguras y ordenadas. El funcionamiento del protocolo se basa en el envío de reconocimientos tras la decodificación de nuevos bloques de información, lo que permite que el destino mantenga una ventana de transmisión deslizante, reduciendo la complejidad del proceso de decodificación. Sin embargo, la sobrecarga de los mensajes de reconocimiento, y la elección de *TCP* como protocolo de transporte puede suponer una pérdida de rendimiento. Kim *et al.* [10] presentaron una evolución de *TCP/NC*, denominada *Network Coded TCP (CTCP)*. En este caso se trabaja a nivel de aplicación, sobre el protocolo *UDP*, aunque se incorporan algunas funcionalidades típicas de *TCP*. Al contrario que *TCP/NC*, la transmisión de bloques de información con un tamaño fijo, ofrece una significativa mejora en el rendimiento, sobre todo sobre redes en las que la presencia de errores de transmisión sea apreciable.

Las soluciones a nivel de transporte no son las únicas que se han propuesto para aumentar el rendimiento sobre redes inalámbricas. También destacan las alternativas a los protocolos de encaminamiento tradicionales. En concreto, las técnicas oportunistas aprovechan el hecho de que todos los nodos puedan escuchar la información transmitida por un ‘vecino’, aunque no vaya dirigida directamente a ellos. Uno de los trabajos más destacados es el de [8], donde se propuso el protocolo *ExOR*, que utilizan como métrica a optimizar el *Expected Number of Transmissions (ETX)* [11], y que necesita una estricta sincronización entre nodos. Precisamente, este es uno de los principales problemas del encaminamiento oportunista; así, el protocolo *MORE* [7] también hace uso de la métrica *ETX* pero, gracias al uso de técnicas de *Network Coding*, elimina la necesidad de dicho sincronismo. Otras propuestas utilizan diferentes métricas, como la de Zorzi *et al.* [12], que usa el *Distance Progress (DP)*, que solo considera la información local disponible en cada nodo, en concreto la distancia hacia el destino; no tiene en cuenta la calidad de los enlaces, que cobra una especial relevancia en redes inalámbricas.

III. IMPLEMENTACIÓN

En esta sección se describe el funcionamiento del esquema propuesto; se presenta en primer lugar las principales características del protocolo *NC intra-flujo* para posteriormente

introducir las diferentes aproximaciones de encaminamiento oportunista que se han empleado.

III-A. Protocolo intra-flujo

Se ha optado por la operación conjunta del protocolo *UDP* con un esquema de codificación aleatoria o *Random Linear Coding (RLC)*. El resultado de la combinación ofrece un servicio similar al ofrecido por *TCP*, ya que la capa de aplicación en el destino recibirá toda la información, de manera ordenada, transmitida originalmente por la fuente. El elemento clave de la solución propuesta es que enmascara la aparición de errores aleatorios durante la transmisión, reaccionando de una forma más apropiada ante las pérdidas que el propio protocolo *TCP*.

Para ello, se ha implementado una entidad *NC* intermedia, que se sitúa entre los niveles de red y de transporte, y cuya tarea principal será la de gestionar todas las operaciones relacionadas con el protocolo desarrollado, entre las que destacan los esquemas de codificación en la fuente, recodificación en los nodos intermedios y decodificación en el destino.

En el nodo origen dicha entidad recibe la información de las capas superiores, almacenando los datagramas *UDP* en un *buffer* de transmisión. Una vez que tiene un bloque completo¹ de K paquetes “nativos”, p_i , la fuente comenzará a generar combinaciones lineales de éstos, dando lugar a una sucesión de paquetes codificados, $p'_j = \sum_{i=0}^{K-1} c_i \cdot p_i$, donde los c_i son escogidos aleatoriamente entre los coeficientes de un cuerpo finito de orden $GF(2^q)$. La fuente seguirá enviando paquetes codificados p'_j hasta que el destino pueda recuperar el contenido original del bloque, situación que notifica a la fuente a través de un mensaje de reconocimiento. Tras la recepción de este *ACK*, el transmisor eliminará los K paquetes nativos que conforman el bloque actual (ya decodificado), pasando a la transmisión del siguiente.

Por otro lado, el extremo receptor hace uso de dos elementos adicionales en el nivel *NC*: por un lado, una matriz C , de dimensión $K \times K$, que mantiene los vectores de coeficientes c_i son transportados en la cabecera del protocolo propuesto. Además incorpora un nuevo *buffer*, con capacidad para alojar K paquetes codificados, en el que almacenará aquellos que contengan información útil para el proceso de decodificación. Así, cada vez que se recibe un paquete, se extrae de la cabecera el vector de coeficientes, insertándolo en la fila r de la matriz C ($r + 1$ es su rango); primero se comprueba si el vector es linealmente independiente de los que se habían almacenado previamente (comprobando si consigue incrementar el rango de C). En caso afirmativo, se considera que el paquete es “novedoso”, por lo que el vector se mantendrá en fila r -ésima de la matriz y el paquete p'_r guardará en el *buffer de decodificación*. Por el contrario, si el vector de coeficientes es combinación lineal de los anteriores, se eliminará de C , ya que la transmisión no ha aportado información adicional. Una vez que la matriz posea rango completo, esto es, haya recibido K paquetes codificados con vectores linealmente independientes, se calculará su inversa, C^{-1} , recuperando los K paquetes originales a través de la expresión $P = C^{-1} \cdot P'$, donde P y P' son las representaciones vectoriales de los

¹En este protocolo se utilizarán bloques de tamaño fijo, cuyo tamaño (K) es un parámetro de configuración del protocolo.

paquetes de datos nativos y codificados, respectivamente. Una vez concluido todo el proceso, se enviará el contenido del bloque a los niveles superiores y, como se ha mencionado anteriormente, se transmitirá un reconocimiento al origen.

Pasando a describir ahora la funcionalidad que implementan los nodos intermedios, hay que distinguir entre dos modos de operación:

1. **Esquema Random Linear Source Coding (RLSC).** En esta primera aproximación, la entidad *NC* no lleva a cabo ningún procesado adicional (“store-and-forward”), por lo que los nodos intermedios simplemente reenvían la información hacia su siguiente salto.
2. **Esquema Random Linear Network Coding (RLNC).** En este caso, se aprovecha la capa de inteligencia que aporta la entidad *NC*, y se les permite transformar el contenido de los paquetes codificados a medida que los procesan. En primer lugar, descartarán todos aquellos que no aporten información. Además, a la hora de reenviar información al siguiente salto, generará nuevas recombinaciones de los paquetes almacenados, $p''_m = \sum_{j=0}^{t-1} c'_j \cdot p'_j$, donde t es el número total de paquetes almacenados en el *buffer*. Estos paquetes p''_m son, a su vez, combinaciones lineales de los K datagramas nativos que forman el bloque original [7].

III-B. Técnicas oportunistas

Como ya se ha adelantado, se combina el esquema de *NC intra-flujo* descrito anteriormente con mecanismos de encaminamiento oportunista. Así, y aprovechando la naturaleza *broadcast* de los canales inalámbricos, aparecerán múltiples rutas alternativas entre dos nodos cualesquiera de una red; en base a esta premisa se intenta alcanzar un equilibrio *óptimo* entre dos aspectos fundamentales:

1. **Redundancia.** Cuando múltiples nodos retransmiten la misma información, se incrementa la robustez del sistema, al aumentar la probabilidad de recibir los mensajes con éxito, limitando la influencia de errores producidos en canales de calidad deficiente. Sin embargo, esto también incrementa la contención del sistema, pudiendo llegar a perjudicar al rendimiento. En ese sentido, merece la pena destacar que la recombinación en los nodos intermedios reducirá el reenvío de información redundante, gracias a la recodificación que llevan a cabo.
2. **Eficiencia.** Se utilizan las rutas que logran maximizar el rendimiento, adaptándose dinámicamente a aquellas que presenten las mejores condiciones.

En la solución propuesta, cada nodo i monitoriza la calidad de los enlaces con cada uno de sus vecinos, en función de la probabilidad de error de trama, $\epsilon_{i,j}$, con $j \neq i$. Para obtener esta información, se utilizará el envío de mensajes o sondas periódicas de señalización (tal y como utilizan varios protocolos de encaminamiento), pudiendo aproximar la calidad de la conexión en función del porcentaje de recepciones correctas en cada uno de los enlaces de interés. En el momento en el que un nodo necesite establecer una ruta hacia un destino, se calculará, a través del grafo subyacente, la métrica *Expected Transmission Count (ETX)* para todos los potenciales candidatos a ser nodos intermedios “activos”.

Antes de comenzar con la descripción del modelo empleado, se presenta a continuación un conjunto de aspectos que se

han considerado para seleccionar la lista de nodos candidatos. En primer lugar, la variable z_i representa el número de transmisiones que un nodo i debe realizar, en promedio, para encaminar un paquete desde el origen hasta el destino; se supone además que los nodos están ordenados según el coste total (*ETX*) hasta el destino. Con estos datos, y utilizando la Ec. 1, se puede calcular el número medio de paquetes que un nodo j ($j \neq i$) espera recibir de nodos con mayor *ETX*, r_j , donde $\epsilon_{i,j}$ se corresponde con la probabilidad de pérdida del enlace $i \rightarrow j$. Se impone la condición de que un nodo j solo reenviará un paquete recibido de un nodo con mayor *ETX* si y sólo si éste no ha sido recibido previamente por otro nodo con menor *ETX*. Por tanto, el número medio de paquetes que el nodo j debe retransmitir, L_j , se puede calcular con la Ec. 2.

$$r_j = \sum_{i>j} z_i (1 - \epsilon_{i,j}) \quad (1)$$

$$L_j = \sum_{i>j} \left[z_i (1 - \epsilon_{i,j}) \prod_{k<j} \epsilon_{i,k} \right] \quad (2)$$

Del mismo modo, se puede calcular el número medio de envíos que un nodo j debe realizar (Ec. 3), teniendo en cuenta que retransmitirá paquetes hasta que un nodo con menor *ETX* lo haya recibido.

$$z_j = \frac{L_j}{(1 - \prod_{k<j} \epsilon_{j,k})} \quad (3)$$

Para llevar la cuenta de las transmisiones en los nodos intermedios, cada uno de ellos manejará un contador interno que se incrementa cada vez que recibe un paquete, utilizando el valor de un parámetro denominado *TxCredit* y que se reduce en una unidad cada vez que se produce un envío. Así, este parámetro refleja el número de transmisiones que un nodo debe realizar por cada paquete que ha recibido desde otro nodo con un mayor *ETX*. Dado que por cada paquete transmitido por la fuente un nodo j recibe $\sum_{j>i} (1 - \epsilon_{j,i}) z_j$, el parámetro *TxCredit* se puede obtener a partir de la Ec. 4.

$$TxCredit = \frac{z_i}{\sum_{j>i} z_j (1 - \epsilon_{j,i})} \quad (4)$$

En concreto, el origen utilizará el algoritmo propuesto en [7] para calcular el *TxCredit* de todos los nodos que participan en la retransmisión de paquetes.

Con ello, el funcionamiento del protocolo es sencillo. El origen utiliza el procedimiento anterior e incluye en la cabecera un vector con los identificadores de los nodos que intervienen en la comunicación y los parámetro *TxCredit* correspondientes. Todos los paquetes serán transmitidos de manera *broadcast*, y los nodos que aparecen en dicha lista serán los encargados de retransmitir el paquete hasta que alcance su destino.

Cada vez que un nodo i recibe un paquete de otro con mayor *ETX* comprueba si se encuentra en la lista, e incrementa su contador con el valor del *TxCredit* correspondiente. Cada vez que el nodo i tiene la oportunidad de transmitir, comprueba si dicho contador es positivo y, si así fuera, transmite un paquete generado como una combinación aleatoria de los anteriormente almacenados, disminuyendo posteriormente el contador en una unidad.

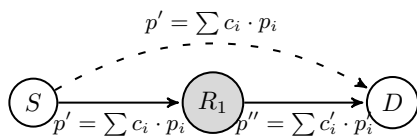


Fig 1: Topología canónica. Recodificación más overhearing

Los paquetes de confirmación utilizados por el protocolo *RLNC* siguen el camino con menor coste, según la métrica *ETX*, entre el nodo destino y origen, utilizando un esquema tradicional de encaminamiento.

IV. SIMULACIÓN Y RESULTADOS

Como se ha dicho anteriormente, el análisis de la solución planteada se realiza inicialmente en un conjunto de topologías canónicas, pues permiten analizar de manera más exhaustiva todo el proceso de codificación/recodificación/decodificación y escucha oportunista. Por ejemplo, en la Figura 1 el nodo origen, S , envía periódicamente paquetes codificados hacia el nodo destino, D . Si se usase un esquema tradicional de transmisión, el nodo R_1 transmitiría los mismos paquetes, p' , que los recibidos a través del enlace $S \rightarrow R_1$ y se podría reducir el número de transmisiones necesarias cuando el nodo D recibiera paquetes por el enlace oportunista, $S \rightarrow D$, y éstos se hubieran perdido por en el enlace convencional, $R_1 \rightarrow D$. En cambio, si el nodo R_1 recombinara los paquetes que tiene almacenados cada vez que transmite un nuevo paquete, aumenta la probabilidad de que la información que viaje por el canal $R_1 \rightarrow D$ sea diferente a la que se transmite por $S \rightarrow D$, siempre y cuando los canales no sean ideales, y no todas las transmisiones llegan a su destino. Este esquema es bastante común cuando el enlace $S \rightarrow D$ tiene una calidad deficiente y, por tanto, el protocolo de encaminamiento establece una ruta alternativa a través con dos saltos con mejores calidades.

Otra posible mejora que surge al combinar Network Coding y escucha oportunista es la posibilidad de que múltiples nodos intermedios participen en la transmisión. En la Figura 2 se muestra un ejemplo ilustrativo, en el que el nodo S transmite información a D , habiendo N nodos intermedios que pueden participar en el reenvío de los paquetes. Así, todos los nodos intermedios escuchan lo que transmite la fuente, retransmitiéndolo al nodo destino. En un caso particular, si la calidad de los enlaces entre el nodo origen y todos los nodos intermedios fuera del 50% (los enlaces entre los nodos intermedios y destino se suponen ideales), la capacidad del canal, bajo un esquema tradicional de encaminamiento, se vería reducida a la mitad. Sin embargo, si se aprovechara la escucha oportunista, se podría reducir la pérdida de paquetes en $(1 - 0.5)^N$, siendo N el número de nodos intermedios, asumiendo una sincronización estricta entre ellos, evitando así retransmitir la misma información, lo que implicaría la utilización de algún protocolo, con la sobrecarga correspondiente. Por otro lado, si se incorpora la capacidad de recodificación a estos nodos intermedios, se evita la retransmisión de información repetida, sin necesidad de utilizar ningún esquema para asegurar la sincronización entre nodos. Si los nodos intermedios transmiten combinaciones lineales de los paquetes que previamente han almacenado, la probabilidad de transmitir información

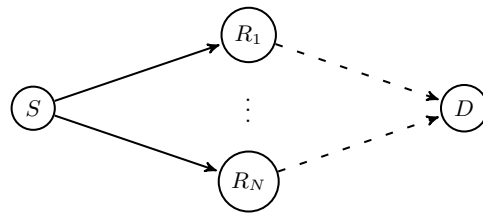


Fig 2: Topología canónica. Recodificación más overhearing

Tabla I: Parámetros de la simulación

Parámetro	Valor
Nivel físico	IEEE 802.11b (11 Mbps)
RETX IEEE 802.11	1 (RLNC), 4 (TCP)
RTS/CTS	Deshabilitado
Tamaño del bloque (K)	32
Orden del cuerpo $GF(2^q)$	$q = 1$
Capa de transporte	UDP (NC)/TCP
Aplicación	$\lceil \frac{10000}{K} \rceil$ paquetes
Tasa binaria (aplicación)	Constante (Saturación)
Payload	1500 bytes en nivel IP
Simulaciones/configuración	50

novedosa al nodo destino incrementa, sobre todo si se trabaja con un cuerpo finito $GF(2^q)$ elevado.

A continuación se presentarán los resultados más relevantes obtenidos a través de una extensa campaña de simulación sobre el simulador `ns-3` [13], que se modificó para integrar la entidad de RLNC, con los procesos de codificación/recodificación/decodificación, tal y como se describen en la sección III. Primeramente se analizará el comportamiento de la solución propuesta sobre las topologías canónicas que se han descrito, para posteriormente extender el análisis a topologías de red más complejas.

En la Tabla I se resumen los parámetros de los escenarios que son comunes a todos los análisis. Como tecnología de transmisión se utiliza IEEE 802.11b a 11 Mbps, con el mecanismo RTS/CTS deshabilitado. Para evitar las colisiones, se establecen dos regiones de cobertura: una inicial, en el que un nodo es capaz de recibir de forma correcta la información (teniendo en cuenta el modelo de error usado), y un área de cobertura mayor, en la que se es capaz únicamente de detectar la transmisión de un paquete. Además, en el caso de RLNC se deshabilitan las retransmisiones a nivel 802.11, al haberse demostrado [14] que influyen negativamente en el rendimiento, mientras que en TCP se mantienen a un valor de 4. A nivel de transporte se utiliza UDP, en el caso de trabajar con RLNC, comparando con el rendimiento observado con la versión TCP "New Reno" [15]. Para evaluar las prestaciones de las diferentes soluciones se estudia el *throughput* que se observa en la entidad receptora, definido como el cociente entre la cantidad de información útil recibida (tamaño del fichero) y el tiempo de transmisión. A nivel de Network Coding se fija un tamaño de bloque, $K = 64$, mientras que el orden del cuerpo finito utilizado se limita a uno binario, es decir $GF(2)$, ya que ofrece un buen ajuste entre rendimiento y complejidad [14] y [16].

Se utilizan dos modelos de errores: en el caso de las topologías sencillas se utiliza una solución simple, estableciendo una *Frame Error Rate* (FER) entre 0.0 y 0.6; es decir, se descartan los paquetes recibidos con una cierta probabilidad, definida por el parámetro FER, siendo cada evento independiente del

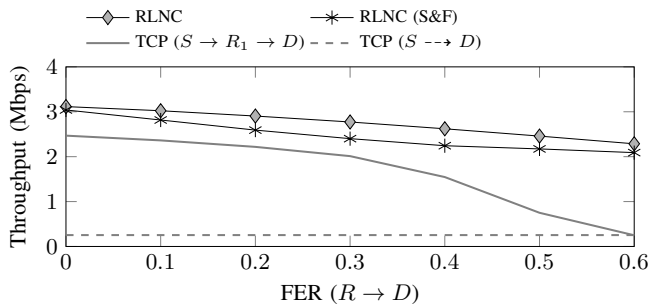


Fig 3: Comparativa del rendimiento del esquema TCP, *RLNC* (*S&F*) and *RLNC*. Escenario en línea. Enlace $R \rightarrow D$ con $FER = 0.0, \dots 0.6$ y un canal directo entre $S \rightarrow D$ con una $FER = 0.6$.

anterior. Sin embargo, teniendo en cuenta que en canales reales (especialmente en escenarios interiores) la pérdida de paquetes es generalmente a ráfagas, también se incorpora un modelo de errores modelado mediante una cadena oculta de Markov (*Hidden Markov Process* (HMP)) [17].

IV-A. Caracterización en topologías sencillas

En primer lugar se utiliza la topología de la Figura 1. En concreto se configura el canal oportunista ($S \rightarrow D$) con una $FER = 0.6$; el canal entre el origen, S , y el nodo intermedio R_1 se supone ideal, mientras que el canal $R_1 \rightarrow D$ se configura con una FER que varía entre 0.0 y 0.6. En la Figura 3 se presenta el rendimiento con dos configuraciones diferentes para el nodo intermedio: actuando como recodificador, *RLNC*, o simplemente reenviando los paquetes recibidos, *RLNC (S&F)*. Es importante destacar que cuando el canal entre el nodo intermedio y el destino es ideal, no existe ninguna ventaja en el del esquema *RLNC*, que incluso genera una ligera pérdida de rendimiento, debido a que la recodificación de paquetes puede aumentar la transmisión de paquetes linealmente dependientes, si el nodo intermedio tiene pocos paquetes en su *buffer* de transmisión. Para calidades peores, la ganancia generada por la recodificación en el nodo intermedio puede llegar al $\approx 16\%$. También se representan los rendimientos observados al utilizar el protocolo TCP, tanto sobre la ruta de dos saltos como sobre el enlace directo $S \rightarrow D$; en cualquier salto, la figura pone de manifiesto la mejora apreciable que el esquema *RLNC* presenta, especialmente al empeorar las condiciones del canal.

Para estudiar las ventajas de la transmisión probabilística, se propone un esquema sencillo, en el que los nodos intermedios transmiten paquetes con una probabilidad establecida a través de la Ecuación 5, en la que el parámetro γ se puede ver como un factor de agresividad: cuanto mayor sea γ , más se favorecerán las transmisiones de los nodos que disponen de mejores enlaces; evidentemente la suma de las probabilidades, para todos los nodos intermedios, es 1. En otro trabajo [18] se ha estudiado con mayor profundidad este esquema, desde una perspectiva más teórica.

$$P_r\{\text{Node } i \text{ forwards packet}\} = \frac{1}{FER_i^\gamma \cdot \sum_{j=1}^N \frac{1}{FER_j^\gamma}} \quad (5)$$

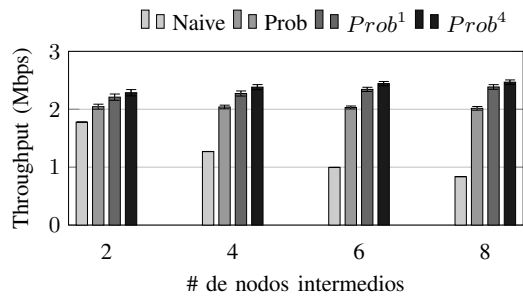


Fig 4: Escenarios múltiples nodos intermedios. Transmisión probabilística

En un segundo análisis, se establece la FER aleatoriamente entre los nodos intermedios y el destino, y se puede comparar el rendimiento de varios esquemas de retransmisión. Una alternativa *naive*, en la que todos los nodos intermedios transmiten información cada vez que tienen la posibilidad de hacerlo (tras recibir un paquete); un primer esquema probabilístico, en el que todos los nodos transmiten con la misma probabilidad, $1/N$, independientemente de la calidad de los enlaces, *Prob*; por último, el esquema regido por la Ecuación 5, $Prob^{\gamma=1}$ o $Prob^{\gamma=4}$. En la Figura 4 se observa que, al incrementar el número de nodos intermedios, el rendimiento observado para el esquema *naive* es cada vez menor, consecuencia del aumento de la contención en el canal. La solución que fija la misma probabilidad de transmitir a todos los nodos mantiene un rendimiento prácticamente constante al aumentar el número de nodos, mejorando de manera sustancial el comportamiento del esquema anterior. Finalmente, si se tiene en cuenta la calidad de los canales en cada uno de los nodos intermedios a la hora de asignar las probabilidades, tal y como hacen los esquemas $Prob^{\gamma=1}$ $Prob^{\gamma=4}$, se pone de manifiesto una ligera mejoría en el rendimiento al incrementar el número de nodos intermedios, debido al aumento de la probabilidad de transmitir por enlaces con mejor calidad. Además, se ve que al utilizar un valor de γ mayor se mejora ligeramente el rendimiento, ya que se favorece el uso de los canales con mejor calidad.

En el escenario anterior, el modelo de error sencillo no añade ninguna correlación en la pérdida de paquetes, de manera que cada trama se pierde con una probabilidad fija e independiente a los eventos anteriores (y posteriores). Con objeto de corroborar que el mecanismo propuesto es viable, se utiliza el canal HMP, configurado para reflejar tres tipos de canal: ideal ($\overline{FER} = 0.0$), bueno ($\overline{FER} \approx 0.163$), intermedio ($\overline{FER} \approx 0.298$) y malo ($\overline{FER} \approx 0.517$).

La idea es comprobar si los paquetes sonda que utilizan varios protocolos de encaminamiento oportunista son apropiados para estimar de manera adecuada la calidad del canal. En concreto se utilizan dos configuraciones, enviando 10 ó 100 paquetes, llamados *Probe Packets*, cada 10 segundos. En la Figura 5 se puede observar, en los ejes superiores, la evolución 'real' del valor de la FER (utilizando el modelo de canal intermedio), y debajo la estimación de la calidad del enlace. Como se puede ver, en ambos casos, Figuras 5a y 5b, se consigue una estimación adecuada de la calidad del canal. A pesar de que la precisión es ligeramente mayor al hacer uso de

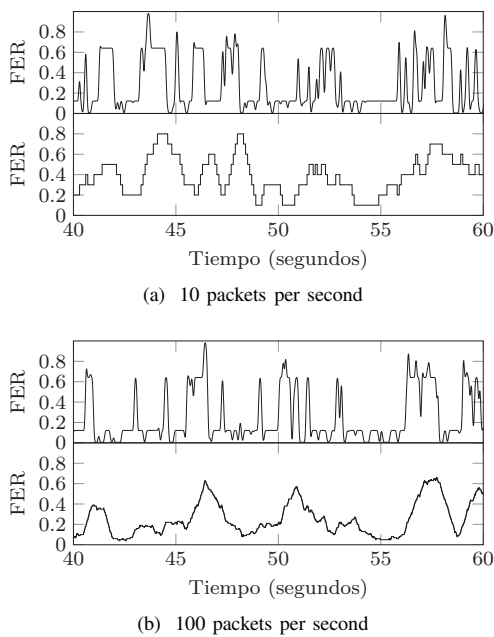


Fig 5: Impacto de la frecuencia de los paquetes sonda sobre la estimación de la calidad del canal

un número mayor de *Probe packets*, hay que tener en cuenta que también se incurre en un aumento de la sobrecarga en la red.

Como último análisis de este primer grupo sobre topologías canónicas, se analiza el rendimiento observado al utilizar las técnicas de transmisión probabilística, pero incorporando las sondas para llevar a cabo la estimación de la calidad de los enlaces, tal y como sucedería en un escenario real. La Figura 6 representa la función de probabilidad acumulada del throughput para tres configuraciones diferentes: un esquema *Naive*, en el que no se utiliza la calidad de los enlaces para establecer la probabilidad de transmisión; el esquema *Probabilístico óptimo*, en el que los nodos son conocedores de la calidad del enlace de manera instantánea, aprovechando la capacidad que ofrece el simulador. Por último, se utiliza el esquema *Probabilístico estimado*, en el que la calidad del enlace se estima tal y como se describió anteriormente. En cualquier caso, y para que la comparativa sea ecuánime, en los tres casos se generan los paquetes sonda según la frecuencia correspondiente. Como se puede ver los dos esquemas probabilísticos consiguen mejorar el rendimiento observado y, aunque el envío de 10 paquetes sonda por segundo reduce ligeramente la ganancia que se consigue con el esquema óptimo, la menor sobrecarga compensa dicha pérdida y el throughput global es mejor en este caso.

IV-B. Extensión a despliegues aleatorios

Posteriormente, se evaluará el rendimiento de las técnicas de encaminamiento oportunista con el esquema RLNC sobre redes inalámbricas malladas. Se utiliza el algoritmo descrito anteriormente para llevar a cabo el encaminamiento y se configuran todos los nodos para que recodifiquen los paquetes que les llegan.

Para evaluar el rendimiento del esquema propuesto en este trabajo y compararlo con una alternativa más tradicional, es

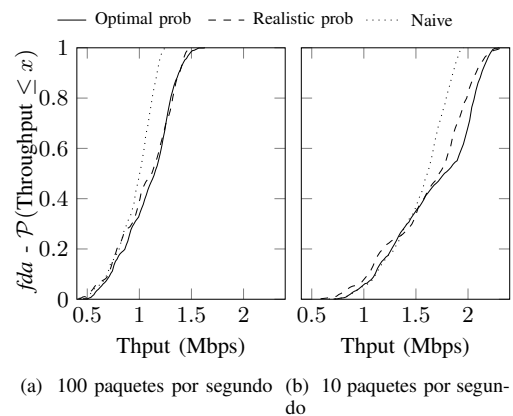


Fig 6: *cdf* del throughput utilizando el esquema probabilístico

necesario generar escenarios aleatorios con un número de nodos. El proceso que se ha seguido es el siguiente:

- Se despliegan 32 nodos de forma aleatoria (proceso de Poisson) sobre una superficie cuadrada de $l \times l m^2$
- La cobertura de cada nodo se modela como un disco de radio de 20 metros. Además se añade un segundo radio, en el que un nodo es capaz de detectar la transmisión sin recibir los datos correctamente, de 40 metros.
- Se descartan todos los escenarios cuyo grafo subyacente no sea conexo; esto es, se asegura que exista una ruta entre cualquier par de nodos de la red.
- Se elijen la pareja origen/destino de forma aleatoria entre los nodos desplegados por la red.

La Figura 7 compara el rendimiento de diferentes esquemas: *RLNC+OR* representa la combinación del encaminamiento oportunista con el esquema de Network Coding. La solución *RLNC* se corresponde con el uso de un esquema tradicional de encaminamiento, en la que los nodos utilizan *NC*; en este caso se utiliza la ruta de menor coste aplicando la métrica ETX. Como se ha hecho hasta ahora, se compara con el rendimiento que se observa al utilizar el protocolo tradicional TCP. Finalmente, para analizar la influencia de la densidad de nodos en la red, se utilizan dos escenarios, con áreas de $100 \times 100 m^2$, y $70 \times 70 m^2$.

Si se observan los resultados de la Figura 7 se ve como el rendimiento del esquema *RLNC+OR* supera a la solución *RLNC*. La ganancia que aporta la solución de encaminamiento oportunista es $\approx 67\%$ en el escenario menos denso y del $\approx 71\%$ en el que la superficie es más pequeña. Esto pone de manifiesto que, al aumentar la densidad de los escenarios, se favorece la aparición de ‘estructuras’ como las que se presentaron en las Figuras 1 y 2, en las que el uso de las técnicas propuestas en este trabajo aportan una ganancia más clara; es por ello que al incrementar la densidad de nodos se aumenta la ganancia en el rendimiento. Ambos esquemas superan de manera clara las prestaciones del esquema tradicional TCP, claramente perjudicadas cuando se incrementan el número de saltos de la ruta o cuando las condiciones de los canales radio empeoran.

Por último, se representa la evolución del rendimiento frente al número de saltos de la ruta con menor coste según la métrica ETX. Se comprueba nuevamente el mal

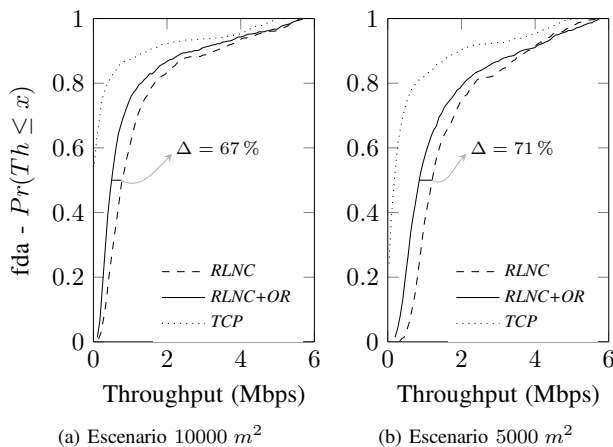


Fig 7: Función de distribución acumulada del throughput

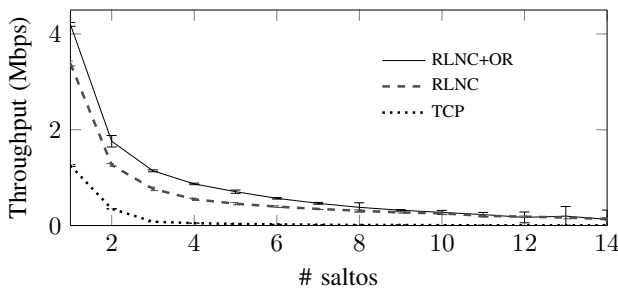


Fig 8: Throughput frente a la número de saltos en la ruta con menor coste

comportamiento del protocolo TCP al empeorar la calidad del canal. Para rutas de un único salto, la mejora del esquema que combina Network Coding y encaminamiento oportunista es cercana a $\approx 3\times$, frente al rendimiento de TCP.

V. CONCLUSIONES Y LÍNEAS FUTURAS

Este trabajo presenta una alternativa para mejorar el pobre comportamiento de TCP sobre redes inalámbricas, ya que reacciona de manera inapropiada ante los errores aleatorios que se producen en medios hostiles. La solución propuesta utiliza como elemento central un esquema de Network Coding intra-flujo, basado en la codificación aleatoria de bloques de tamaño fijo. Sobre esta solución se plantean diferentes alternativas para establecer un encaminamiento oportunista que, además de transportar de manera eficiente la información utilizando múltiples rutas simultáneas entre origen y destino, mejoran la robustez frente a las pérdidas, maximizando al mismo tiempo la utilización del canal. A pesar de que se utilice UDP como protocolo de transporte, el servicio de la solución planteada es equivalente al que ofrece tradicionalmente TCP.

Para evaluar su rendimiento, comparándolo con una solución clásica basada en TCP, se ha llevado a cabo una profunda campaña de simulación sobre la plataforma ns-3, en la que se han estudiado sus prestaciones, tanto sobre topologías canónicas como sobre despliegues aleatorios. Los resultados han puesto de manifiesto una mejora notable, caracterizando el throughput medido en la aplicación del destino. La ganancia se ve acentuada a medida que la calidad de los

enlaces inalámbricos empeora, demostrando la capacidad de la solución propuesta para enmascarar los errores aleatorios producidos en el medio físico.

Como principal línea de investigación abierta tras la realización de este trabajo, se puede destacar la posibilidad de utilizar el marco desarrollado para analizar diferentes esquemas, tanto de Network Coding, incluyendo el análisis de la integración de técnicas que combinan paquetes pertenecientes a diferentes flujos (Inter-Flujo), como de encaminamiento oportunista. En este último grupo se tratará de utilizar el tráfico de datos para llevar a cabo la monitorización de la calidad de los enlaces, evitando incurrir en la sobrecarga generada por los paquetes sonda.

AGRADECIMIENTOS

Los autores querían expresar su agradecimiento al Gobierno de España por su financiación en el proyecto “Connectivity as a Service: Access for the Internet of the Future”, COSAIF (TEC2012-38574-C02-02).

REFERENCIAS

- [1] I. F. Akyildiz, X. Wang, and W. Wang, “Wireless mesh networks: a survey,” *Computer Networks*, vol. 47, no. 4, pp. 445 – 487, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128604003457>
- [2] F. Lefevre and G. Vivier, “Understanding TCP’s behavior over wireless links,” in *Communications and Vehicular Technology, 2000. SCVT-200. Symposium on*, 2000, pp. 123–130.
- [3] M. Zorzi, A. Chockalingam, and R. Rao, “Throughput analysis of TCP on channels with memory,” *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 7, pp. 1289–1300, July 2000.
- [4] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, “XORs in the Air: Practical Wireless Network Coding,” *Networking, IEEE/ACM Transactions on*, vol. 16, no. 3, pp. 497 –510, June 2008.
- [5] F. Zhao, M. Medard, M. Hundeboll, J. Ledet-Pedersen, S. Rein, and F. Fitzek, “Comparison of analytical and measured performance results on network coding in IEEE 802.11 ad-hoc networks,” in *Network Coding (NetCod), 2012 International Symposium on*, June 2012, pp. 43–48.
- [6] D. Gomez, S. Hassayoun, A. Herren, R. Aguero, and D. Ros, “Impact of network coding on tcp performance in wireless mesh networks,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, Sept 2012, pp. 777–782.
- [7] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, “Trading structure for randomness in wireless opportunistic routing,” in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM ’07. ACM, 2007, pp. 169–180. [Online]. Available: <http://doi.acm.org/10.1145/1282380.1282400>
- [8] S. Biswas and R. Morris, “Opportunistic routing in multi-hop wireless networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 69–74, Jan. 2004. [Online]. Available: <http://doi.acm.org/10.1145/972374.972387>
- [9] J. Sundararajan, D. Shah, M. Médard, S. Jakubczak, M. Mitzenmacher, and J. Barros, “Network Coding Meets TCP: Theory and Implementation,” *Proceedings of the IEEE*, vol. 99, no. 3, pp. 490 –512, March 2011.
- [10] M. Kim, J. Cloud, A. ParandehGheibi, L. Urbina, K. Fouli, D. J. Leith, and M. Médard, “Network Coded TCP (CTCP),” *CoRR*, vol. abs/1212.2291, 2012.
- [11] D. S. J. de Couto, D. Aguayo, J. Bicket, and R. Morris, “A high-throughput path metric for multi-hop wireless routing,” in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’03. New York, NY, USA: ACM, 2003, pp. 134–146. [Online]. Available: <http://doi.acm.org/10.1145/938985.939000>
- [12] M. Zorzi and R. Rao, “Geographic random forwarding (geraf) for ad hoc and sensor networks: multihop performance,” *Mobile Computing, IEEE Transactions on*, vol. 2, no. 4, pp. 337–348, Oct 2003.
- [13] “The ns-3 network simulator,” 2008, <http://www.nsnam.org/>.
- [14] D. Gomez, E. Rodriguez, R. Aguero, and L. Munoz, “Reliable communications over lossy wireless channels by means of the combination of udp and random linear coding,” in *Computers and Communication (ISCC), 2014 IEEE Symposium on*, June 2014, pp. 1–6.

- [15] S. Floyd, A. Gurtov, and T. Henderson, "The newreno modification to tcp's fast recovery algorithm," 2004.
- [16] P. Garrido, D. Gomez, J. Lanza, R. Aguero, and J. Serrat, "Análisis de las prestaciones de un esquema de codificación lineal aleatoria sobre redes inalámbricas," in *XII Jornadas de Ingeniería Telemática (JITEL 2015)*, Universidad of Mallorca, October 2015, pp. 1–1.
- [17] D. Gómez, R. Agüero, M. García-Arranz, and L. Muñoz, "On the use of hidden markov processes and auto-regressive filters to incorporate indoor bursty wireless channels into network simulation platforms," *Wireless Networks*, pp. 1–18.
- [18] P. Garrido, D. Gomez, R. Aguero, and L. Munoz, "Performance of random linear coding over multiple error-prone wireless links," *Communications Letters, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

Performance evaluation of dissemination protocols for emergency messages in Vehicular Ad-Hoc networks

Cristhian Iza Paredes, Ahmad Mohamad Mezher, Mónica Aguilar Igartua
Department of Network Engineering,
Universitat Politècnica de Catalunya (UPC)
C/ Jordi Girona 1-3, 08034 Barcelona, Spain.
[cristhian.iza, ahmad.mezher, monica.aguilar]@entel.upc.edu

Abstract—Road safety applications envisaged for VANETs depend largely on the exchange of messages to deliver information to the concerned vehicles. In the last years, several dissemination protocols have been presented in the literature. In this paper we study and evaluate the dissemination of emergency messages in realistic vehicular scenarios. In the evaluation, some protocols are examined from two perspectives: qualitative and quantitative. We identify the principal mechanisms of dissemination and examine the factors that most impact on the simulation results. In addition, we investigate the effects of the shadows of buildings and other vehicles in the performance of the dissemination protocols. The simulation results suggest the need to include scenarios with fixed and mobile obstacles to increase the credibility of the performance of the evaluated protocols.

Keywords—VANET, V2V, VEINS, OMNET++, SUMO

I. INTRODUCTION

According to the World Health Organization (WHO), traffic accidents are the leading cause of death in young people between 5-29 and the eighth leading cause of death in the world. Therefore, traffic accidents have become a major public health problem worldwide. Intelligent Transportation Systems (ITS) promise to reduce the number of accidents, their severity, and the time spent by the emergency services to attend a traffic accident [1]. ITSs integrate an extensive range of wireless and wired communications, information processing, advanced computing, and electronic technologies. One of these technologies is an important line research supported by many projects around the world dedicated to the study of Vehicular Ad hoc NETWORKS (VANETs) [2].

VANETs are a type of mobile ad-hoc networks wherein the network topology changes dynamically due to the high mobility of vehicles. The most promising application of VANET is road safety. Several operational tests have demonstrated the feasibility and effectiveness of this service. However, there are still many challenges to integrate wireless communication between vehicles. These include both engineering methods as well as the design of communication protocols.

Our area of research focuses on the evaluation of dissemination multi hop protocols in VANETs. In this paper, we first evaluate the performance of several dissemination schemes. Our work gives a qualitative assessment of the

analysed protocols. For this, first we have summarized the main schemes used in dissemination of messages. Second, a quantitative evaluation of dissemination protocols are presented. We have selected some of the most representative dissemination protocols proposed in the current literature. In our evaluation we have used realistic scenarios which take into account the factors that most impact on the performance evaluation. A comparative results of the quantitative assessment based on traditional evaluation metrics are presented, and we include a special metric that reflects an overall performance of each protocol.

The rest of the paper is organized as follows: Section II presents the problem statement. Section III describes the qualitative analysis of the protocols. Afterwards, Section IV discusses the performance evaluation and includes the results of our analysis. Finally, Section V presents conclusions and future work.

II. BACKGROUND

A. Main schemes

All road safety applications assume that warning messages come from an infrastructure or from the vehicles themselves. The broadcast message typically refers to the process of spreading data over distributed wireless networks. If we analyze the dissemination of messages from the networking point of view, it requires broadcast capabilities at the link layer, allowing a frame to be transmitted to all the vehicles in the transmission range. Message broadcast also assumes implementation of network mechanisms to disseminate the message in the whole network. In this context, a message dissemination process can be analyzed under three kinds of schemes: flooding, relaying and opportunistic forwarding.

First, a traditional approach to disseminate messages in vehicular networks is Flooding. Once a message is generated and transmitted by a vehicle, it is relayed once by all the receivers (i.e., vehicles within the sender's transmission range). This approach can provide a very high speed message dissemination. However, due to the shared wireless medium, Flooding can also lead to frequent collisions in the transmission of messages specially in vehicular areas with high density. This problem is referred to as *broadcast storm*. Some solutions that

seek to alleviate the problem of broadcast storms in mobile ad-hoc networks were proposed in [3].

The second scheme is called Relaying. This technique consists on selecting a subset of vehicles which are located along the road to act as relaying nodes. The goal here is to select relaying vehicles found in preferred positions, while other vehicles are inhibited for retransmission. Some protocols include the information provided by digital maps and global positioning systems (GPS) for the selection of relaying nodes. When we consider the space environment, a vehicle can determine the most appropriate neighbor. Recently, in [4] authors suggest the size of the vehicle as an additional variable for selecting the relaying vehicle. According to the published results, high vehicles have in average a larger communication range which minimizes the number of hops.

Finally, Opportunistic Forwarding is a scheme that can be useful when network fragmentation occurs in periods of low traffic density. This mechanism is also known as store-carry-forward (SCF). It allows the vehicle to store in a buffer while the vehicle looks for another vehicle to forward the message. Many protocols use this approach in environments where it is not possible to ensure a continuous network connectivity.

B. Fair evaluation in realistic scenarios

The deployment and testing of VANETs involves high investments, and in most cases it is economically prohibitively. In the academic community, communication protocols among vehicles are usually evaluated through simulation techniques. Many advances have been made toward making realistic simulations, but still there are factors that are not taken into account in the simulations. According to the authors in [5], five factors have a strong influence on the quality of performance evaluation of protocols in vehicular environments: a realistic mobility pattern of vehicles in the simulation, a realistic scenario, realistic propagation models of the radio signal, appropriate evaluation metrics and realistic human driver behavior.

In this work, we selected VEINS [6] to carry out our simulations. VEINS is a simulation framework that couples the real-time network simulator OMNeT++ [7] with the mobility generator SUMO [8]. VEINS has important features such as simulation framework for vehicular networks, online reconfiguration and re-routing of vehicles, fully detailed models of the IEEE 802.11p standard, IEEE 1609.4 WAVE [9], support for realistic maps and realistic traffic. SUMO reports to OMNeT++ on new vehicles and their current positions in each simulation step using Traci interface. In turn, OMNeT++ creates instances of nodes in the network simulation so that they can communicate. SUMO is capable of assuming a different motion to each vehicle to attain scenarios close to real environments.

The selection of the simulation scenario has a great influence on the results of a performance evaluation of vehicular communication protocols. The first step towards the definition of the scenario is the selection of the roads. In order to be as much realistic as possible, we use real maps extracted from OpenStreetMap [10] to prepare scenarios with highways and roads in urban environments. One of the biggest challenges facing VANET networks in urban scenarios is the presence of buildings and other artificial structures that affect wireless connectivity. It is very important that obstacles can be modeled

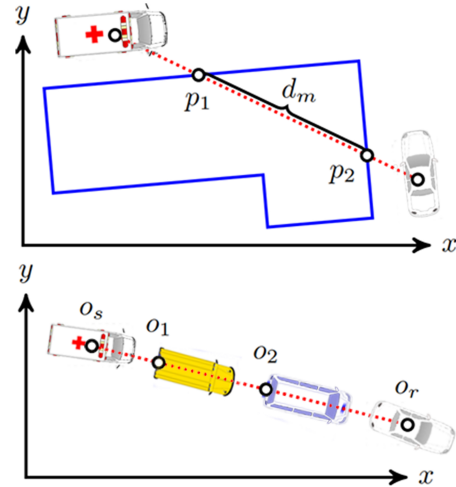


Figure 1: Calculation of line of sight intersection points with building and vehicles [12]

in the simulation to obtain accurate results, and to evaluate how the protocol takes special measures to overcome the building issue. Shadows in radio communication produced by buildings is an important factor for road safety critical applications. In [11], authors presented a computationally affordable simulation model for obstacles with IEEE 802.11p in urban environments. It is an empirical model based on measurements of the real world, and it only considers the line of sight between transmitter and receiver. This model of obstacles is integrated with the framework of VEINS. Commonly available geodata is used to model buildings and the respective radio signal shadowing in vehicular network simulation; for instance, OpenStreetMap provides this kind of information.

The impact of shadows on the radio communication caused by vehicles was evaluated in [13]. The aim of the study was to show that the impact of the blockage of vehicles on the radiocommunications is not negligible. Based on this model, algorithm depicted in Fig. 2 shows how we have added an additional calculation of the attenuation due to vehicles in the framework of VEINS. To compute the impact of moving vehicles on the power loss, we employ a similar technique to presented in [13] and [12]. The main purpose is to identify the group of vehicles (o_1, o_2, \dots, o_n) that intersect the direct line of sight between two communicating vehicles o_s and o_r as it is shown in Fig. 1. According to ITU-R recommendations [14], the signal power loss ($L[dB]$) caused by an obstacle between a sender and a receiver can be calculated using a *single knife edge* approximation. To make such calculations it is necessary to idealize the form of the obstacles, all the geometrical parameters are combined together in a single dimensionless parameter normally denoted by v . It is a geometrical parameter used to determinate how much of the *first Fresnel zone* is obstructed by the obstacle.

$$v = h \sqrt{\frac{2}{\lambda} \left(\frac{1}{d_1} + \frac{1}{d_2} \right)} \quad (1)$$

where h is the difference between the height of the obstacle and the height of the straight line joining sender and receiver,

```

ATTENUATION( $x, y$ )
1: for  $i = 1$  to  $r$  do
2:    $[coord] = getIntersect(i)$ 
3:   if SIZE( $[coord]$ )  $\neq 0$  then
4:      $att = calcAddAtten([coord])$ 
5:   else
6:      $att = 0$  dB
7: return success
    
```

Figure 2: Algorithm to compute the additional attenuation due to vehicles [13]

d_1 and d_2 are the distances from sender and receiver to the obstacle, respectively. We use the appropriate wavelength ($\lambda=0.05$ m) for the standard for VANET communication which operates in the 5.89 GHz frequency band.

In the implemented model in [14], power loss is assumed to only occur for a geometrical parameter of $v > -0,7$. In this case the loss is calculated by:

$$L[dB] = \begin{cases} 6.9 + 20\log_{10}(\sqrt{(v - 0.1)^2 + 1} + v - 0.1) & \text{for } v > -0,7 \\ 0 & \text{; otherwise} \end{cases} \quad (2)$$

This simple *single knife edge* method can be generalized to multiple obstacles.

On the other hand, one of the major problems when one wants to compare the performance of dissemination protocols is that researchers often use different metrics in assessing the performance of their protocols. In fact, they usually present the metrics which are advantageous to the proposed protocols, while other metrics are ignored. Thus, based on the results reported in the original documents, it is not possible to make a quantitative comparison among the different proposal effectively. Therefore, besides using conventional metrics, we assess the performance of protocols based on the unified metric presented in [15]. This metric is called Dissemination Efficiency (DE) and it is a measure that includes three separate domains (frequency, time and space) as it can be seen in Table I. The DE metric is defined as:

$$DE = \frac{\text{Propagation Distance} \times \text{Success Rate}}{\text{Propagation Time} \times \text{Redundancy Rate}} \quad (3)$$

Intuitively, DE measures how far an information packet can propagate through the network per unit of time and per amount of overhead generated. The propagation distance is measured in meters, the propagation time is measured in seconds, the redundancy rate and the success rate are unit-less; therefore, DE has a unit of m/s.

Finally, in most cases technical factors are not sufficient to globally evaluate a protocol for vehicular communication. A key component which is not very often found in the evaluation of protocols or applications traffic is the interaction of the system with human behavior, i.e. the driver. Although this problem can be critical for accurate and realistic assessment of performance, in this paper, we have assumed that drivers act exactly as expected or as suggested by the road safety application.

Domain	Metric	Description	Favorable Value
Frequency	Redundancy Rate	Measures the number duplicate packets per one source packet	Low
	Success Rate	Measures the proportion of vehicles that successfully receive the broadcast packets	High
Space	Propagation Distance	Measures the distance between the origin of the packet and the point where it is last received	High
Time	Propagation Time	Measures the time it takes a packet traverse from a source to a specific point in the network	Low

Table I: Metrics to calculate Dissemination Efficiency.

III. QUALITATIVE ASSESSMENT

In this section some of the protocols to disseminate messages are analysed and compared qualitatively. Table II presents a qualitative comparison of solutions for message dissemination in VANETs. The comparison is based on three sets of criteria: *forwarding strategy*, *scenarios and assumptions*. The forwarding strategy defines schemas used to route messages toward the destination. The scenario approach categorizes protocols based on operating environments of each solution. Assumptions identify external sources of information used by each protocol for operation.

One of the main challenges in the dissemination of messages is to mitigate the broadcast storm in vehicular environments with high density of vehicles e.g., traffic jams. In [16], authors presented three techniques known as Weighted p-Persistence, Slotted 1-Persistence and p-Persistence. The overall objective of these solutions is to assign the highest relaying priority to the most distant vehicles in the direction of the message. In the Weighted p-Persistence scheme, the highest probability of retransmission is assigned to the farthest vehicles inside the communication range of the transmitter vehicle. In the Slotted 1-Persistence scheme, vehicles are allocated in different time intervals to wait before retransmission. Each time interval is calculated according to the number of intervals, the distance between vehicles (transmitter - receiver) and the transmission range. The farthest vehicles from the vehicle sender of the message have the highest priority and the shortest possible waiting time before retransmission. Slotted p-Persistence mixes probability and delay by giving vehicles with the highest priority the shortest delay and highest probability to rebroadcast. This would be an indication that the information has already been disseminated and redundant rebroadcasts can be suppressed. The proposed techniques are distributed and relay on GPS information, but they do not require any other prior knowledge about network topology.

Several dissemination protocols usually require external information about the network topology to select the vehicles to relay messages. For example, in [17] authors presented the protocol called DV-CAST. This protocol mitigates the broadcast storm and network partitions. The protocol uses the local topology information (list of neighboring vehicles) as the main criterion for message relay. The diffusion process is adapted according to the density of neighboring vehicles, its position and direction. If the local topology information has a high density of vehicles, DV-CAST applies the suppression of

Protocols V2V	Forwarding strategy						Scenario				Assumptions			
	Position based	Counter based	Probabilistic based	Distance based	Local topology based	Delay Based	Tall Based	Store carry forward	Highway	Urban	GPS	Map	Adaptive beaconing	Neighbors info
Slotted p-Persistence	✓		✓	✓		✓					✓			
Slotted l-Persistence	✓			✓		✓					✓			
Weighted p-Persistence	✓		✓	✓		✓					✓			
UV-CAST	✓			✓	✓	✓		✓		✓	✓			✓
DV-CAST	✓		✓	✓	✓	✓		✓	✓		✓			✓
AMD	✓			✓	✓	✓		✓		✓	✓			✓
NSF	✓					✓		✓		✓	✓			✓
NJL	✓					✓		✓		✓	✓			✓
RTAD	✓	✓				✓		✓		✓	✓			✓
RCP	✓			✓		✓		✓		✓	✓			✓
APAL	✓	✓	✓			✓		✓		✓	✓			✓
NoG	✓					✓		✓		✓	✓		✓	✓

Table II: Classification of message dissemination protocols.

broadcast (Weighted p-Persistence). Conversely, if the vehicle density is low, this protocol uses the opportunistic forwarding mechanism known as Store-Carry-Forward (SCF). This mechanism can take advantage of the mobility of vehicles to store and forward messages when vehicles are separated geographically.

Due to possible changes in direction of vehicles at intersections, it is not obvious which vehicles should be responsible for forwarding the warning message. Unlike highway scenarios, where the temporal relay node is always the farthest vehicle traveling in the direction opposite to the message direction, the same method may not work in urban scenarios. Thus, protocols should take into account the challenges of dissemination in urban environments. For example, authors proposed a protocol called UV-CAST for urban areas in [18]. The protocol uses digital map information to verify if it is at an intersection or not. The waiting time to relay the message is less at intersections. Once the timer expires the vehicle relays the message only if it does not receive duplicate messages, otherwise the relay is inhibited. Additionally, UV-CAST can assign to more than one vehicle the responsibility for opportunistic forwarding (SCF), so these vehicles can forward the message more than once.

Despite numerous dissemination protocols in the literature, there are only a few protocols developed to scale properly in various network densities in both realistic highway and urban scenarios. For example, authors in [19] presented a protocol called AMD (Adaptive Multi-directional data Dissemination). The protocol disseminates the message to multiple addresses, which are adjusted adaptively according to the local map and the GPS information. DV-CAST [17], UV-CAST [18] and AMD [19] handle a similar scheme that combines broadcast suppression and store-carry-forward technique, i.e. these proposals tackle the broadcast storm and the disconnected network problems simultaneously. According to the good results, this combination is an important basis for the development of a protocol for road safety applications.

Some dissemination protocols also adapt to the vehicles' density for message dissemination strategy. Indeed, in [20] authors presented two approaches for dissemination: NFS (Neighbor Store and Forward), a protocol for scenarios with low traffic density; and NJL (Nearest Junction Located), a scheme for vehicular scenarios with high density. Both protocols maintain a list of neighbors which is constructed by

exchanging beacon messages. NSL is designed to relay the message only if the vehicle is the closest to any intersection. NSF is based on the mechanism of opportunistic forwarding.

The region of interest (ROI) for disseminating emergency messages in urban area is deepened in [21]. They propose the protocol called CPR (Road Casting Protocol). It is designed for sending emergency messages to a group of vehicles identified by the road segment on which they are located. Each emergency message receiver decides to forward the message based on the incident and the receiver's location relative to a point called Critical Junction (CJ). This point is an intersection beyond which a vehicle can not avoid the blocked road segment. To select the vehicle to forward the message, the protocol is based on two factors: distance and link quality. The distance factor considers the position of vehicles (receiver - transmitter) and the next intersection. The link quality factor takes into account the signal quality, the channel quality and the probability of packet collision.

So far, all proposals assume the availability of global positioning system. However, some authors present solutions that do not use the information provided by the GPS. For example, in [22] authors presented a protocol called APAL (Adaptive Probability Alert Protocol). This protocol uses adaptive probability and time intervals to trigger retransmission. Another dissemination protocol that avoids the use of GPS is proposed in [23] where the NoG (Nongeographical Knowledge Broadcasting Protocol) is presented. This interesting proposal based on graph theory, consists of three main modules: A beacon mechanism, a broadcast mechanism and a waiting time mechanism. These latter mechanisms depend on the accuracy of the information provided by the beacons. For this, the authors incorporate a mechanism to adapt beacon frequency to the vehicular density.

In Table II, we summarize the main characteristics of each analysed protocol. It can be noted that all these protocols basically use the same mechanisms for selecting nodes to forward the message: position, counter, distance, probability, waiting times, local topology and store-carry-forward. These protocols are mainly focused on reducing latency and the mitigation of broadcast storms. However, the results are limited to simulations with unrealistic scenarios where all protocols achieve high performance. We consider that these protocols should be evaluated under more realistic scenarios to reach a

fair assessment. That has been our goal in this work.

IV. QUANTITATIVE ASSESSMENT

In this section a performance assessment of dissemination schemes is carried out by means of simulations. Our goal is to study the dissemination of emergency messages under urban realistic scenarios. We first present the simulation setup used including models and scenarios. Then, we analyse the dissemination of messages applied by several selected protocols.

A. Simulation setup

To carry out the performance of the analysed dissemination schemes, we use VEINS [6]. This framework builds on the MiXiM framework physical layer model, which allows the implementation in the simulator of the building and vehicle shadowing models discussed in Section II-B. We have provided each run with a different random scenario that fulfills the requirements of the study. For each point in all figures we have calculated the average from 10 simulation runs. This let us obtain a standard error less than 5% in a 95% confidence interval. The packet error and Medium Access Control (MAC) layer models adopted are based on the IEEE 802.11p, using a data rate of 6 Mbit/s, a transmission power of 20 mW, and a receiver sensitivity of -94 dBm. For eliminating effects caused by switching channel between the Control Channel (CCH) and the Service Channel (SCH), we changed the model to use only the CCH. In addition, all beacons use the same Access Category (AC_BE), which results in the Contention Window (CW) and AIFSN parameters mentioned in Table III.

	Parameter	Value
Physics and Mac Layers IEEE 802.11p	Channel	Channel 178, 5.89 GHZ
	Bandwidth	10 MHz
	Transmission range	230m
	Transmission power	20 mW
	Sensitivity	-94 dBm
	Obstacle model	Defined in [13], [12]
	CWmin, CWmax	15,1023
	AIFSN	2
	Bit rate	6Mbit/s
Broadcast Supression mechanism	τ	5ms
	N_s	3
	tmax	500ms
	Beacon frequency	1 Hz
	Beacon size	>=32 bytes
Scenarios	Data Message size	2312 bytes
	Number of Runs	10
	Time to live (TTL)	90s
	Density	20 - 300 (veh/km2)

Table III: Simulation parameters.

For all simulation scenarios, the data message size is 2,312 bytes, i.e. the maximum allowed by the IEEE 802.11p standard. This allows us to evaluate the protocols in the worst-case scenario in terms of medium occupation caused by the transmission of messages. Beacons are sent at the frequency of 1 Hz. This is usually the highest frequency expected to be used for the transmission of beacons which gives the worst-case scenario in terms of freshness of the one-hop neighborhood information. In order to study realistic vehicle-caused radio shadowing, we used a typical mix of different vehicles (90% cars and 10% trucks). Only trucks can attenuate or even

block the signal generated from cars. All vehicles are moving according to the SUMO standard Krauss driver model.

B. Scenario Description

We focus on the immediate consequences of an accident. The crashed vehicle starts to generate and transmit an emergency message after the collision to inform neighboring vehicles as quickly as possible in a distributed way. In the simulations, we used a real city area obtained from Barcelona, Spain as our urban scenario (see Fig. 3). This segment has an area of 1.5 x 2km² and was retrieved from OpenStreetMaps [10]. A vehicle positioned approximately at the center of the network is responsible for generating a single message to be disseminated in a TTL of 90s.

Shadowing models are used to reproduce the attenuation of a radio signal induced by obstacles, such as buildings or other vehicles blocking the direct line of sight. Figures 3 and 4 show the map section considered, where buildings represented by pink rectangles are radio obstacles. For all the experiments, the results for the three types of shading models are presented: Clearances (F), shadow Building (B), shadow Buildings + Shadow vehicles (B+V):

Clearances (F): Assuming free-space propagation, path loss is estimated by taking the distance to the receiver and the wavelength.

Shadow Building (B): Shadowing caused by buildings. The main idea is to count the number of exterior walls of a building to approximate the impact of the radio-signal shadowing caused by exactly this building.

Shadow Buildings + Shadow vehicles (B+V): The impact of radio-signal shadowing caused by buildings and other vehicles.

C. Metrics

We use four metrics to evaluate each message dissemination protocol: packet delivery ratio, average delivery ratio, total transmitted messages and efficiency dissemination.

Packet Delivery Ratio: It indicates the percentage of nodes that receive a single emergency message within a specified period.

Average Packet Delay: It provides the indication of how soon the message can be delivered to the intended receiver. This is an important metric for emergency messages that must be disseminated as rapidly as possible.

Transmitted Messages: It is the total number of data messages disseminated by all vehicles in the network.

Efficiency Dissemination. It measures how far an information packet can propagate through the network per unit time and per amount of overheads generated.

D. Simulation Results

In this section we present some representative simulation results. We select two state-of-the-art protocols for comparison, called: UV-CAST [18] and Slotted 1-Persistence [16]. Although this comparison might seem unfair because of the features of each protocol, this work attempts prepare a testing



Figure 3: Urban scenario: a 1.5x2 km² region of the city of Barcelona, Spain.

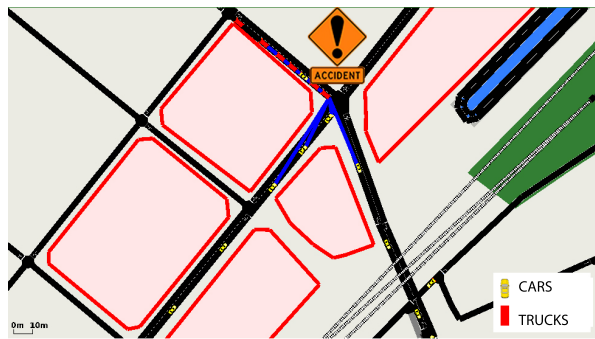


Figure 4: Traffic in urban areas with shadows on the radio communication.

platform to evaluate message dissemination protocols in realistic scenarios. To do so, we have implemented the code of both protocols in the VEINS simulation framework.

Slotted 1-Persistence is a mechanism of suppression of broadcast storms. The broadcast coverage is spatially divided into regions (slots), and a shorter waiting time is assigned to the nodes located in the farthest region. Each node uses the GPS information to calculate the waiting time to retransmit. To alleviate the broadcast storm this scheme inhibits retransmissions in some vehicles to reduce redundancy and therefore the contention and collisions. We define a number of slots $N_s=3$. UV-CAST is a protocol that specifically addresses urban scenarios. It combines a suppression technique for dense networks that gives higher priority to vehicles near intersection points and a mechanisms to select vehicles to store, carry, and forward packets. The vehicular density varies from 20 to 300 vehicles/km².

In a first set of experiments, we evaluated the performance of the packet delivery ratio. Fig. 5 shows the packet delivery rate for three radio shadowing models. In the case of free space (F), UV-CAST achieves successful delivery rates close to 100% above 100 vehicles/km². The protocol uses only a subset of vehicles the task of opportunistic forwarding (store-carry-forward). Thus, uninformed vehicles that do not find a vehicle from this subset will not receive the disseminated message. It is clear the poor performance of Slotted-1-Persistence for low traffic density (3 inferior lines in Fig. 5). This result is expected, because this protocol was designed for high density scenarios. However, Slotted-1-Persistence only reaches delivery rates under 80%. As this protocol does not have a mechanism for opportunistic forwarding, dynamic topology

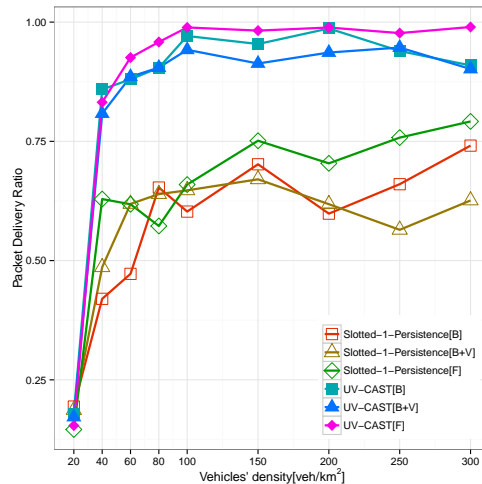


Figure 5: Packet delivery ratio with 95% confidence intervals for different network densities in an urban scenario.

networks in VANET causes temporary disconnections, interrupting the dissemination and compromising the delivery of the messages. Additionally, the results presented in Fig. 5 corroborate the impact that buildings and high vehicles have as obstacles in the line of sight in both protocols. There is a difference on average of 3% between packet delivery rate in free space (F) and delivery rate with shadows of buildings (B). This difference achieves on average 7% if building and high vehicles (B+V) are considered. Although we expected

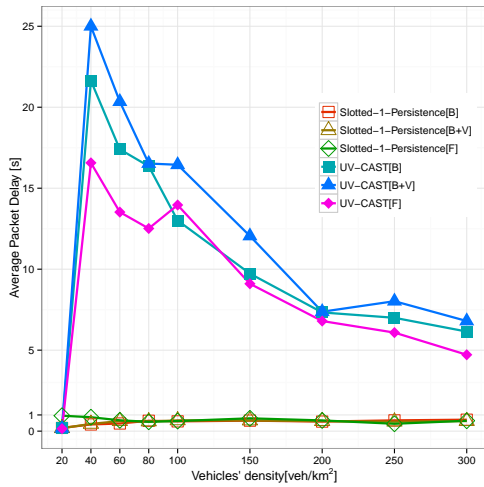


Figure 6: Average packet delay with 95% confidence intervals for different network densities in an urban scenario.

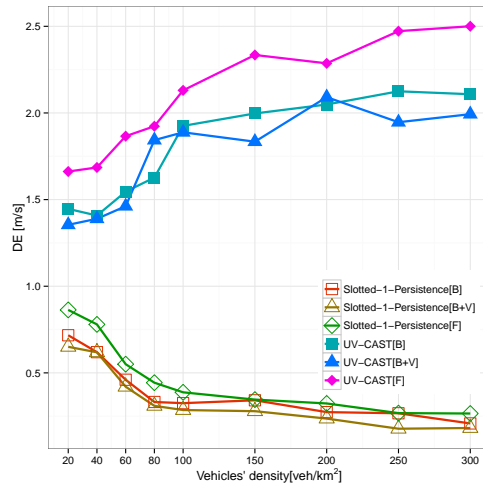


Figure 8: Dissemination Efficiency with 95% confidence intervals for different network densities in an urban scenario.

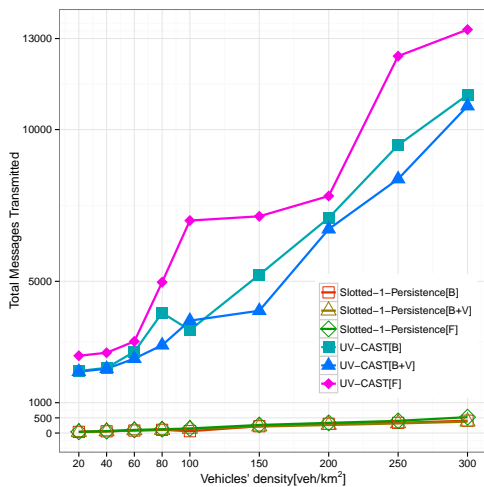


Figure 7: Total messages transmitted with 95% confidence intervals for different network densities in an urban scenario.

a more noticeable difference, we believe that the ability of the dissemination protocols conceals a greater impact of the shadows from buildings and high vehicles.

In a second step, we investigate the performance of latency. Fig. 6 shows the average packet delay to deliver the message for the three radio shadowing models. In the case free space (F), the lower delay (lower than 1s) for UV-CAST when the traffic is 20 vehicles/km² is because it fails to deliver the message to almost all vehicles of the region of interest as it was shown in Fig. 5. The higher delays for UVCAST under low traffic densities (i.e., equal or lower than 150 vehicles/km²) are explained by the store-carry-forward mechanism used by the protocol. As the traffic density increases, the delay decreases, since the message can be rapidly disseminated through direct relaying, thus requiring less the use of the store-carry-forward mechanism. Similarly, the results presented in Fig. 6 confirm the impact that buildings and high vehicles have as obstacles in the line of sight in UV-CAST protocol. There is a

difference between average packet delay in free space (F) and delivery rate with shadows of buildings (B). This difference is more pronounced if building and high vehicles (B+V) are considered. If the scenario is densely congested, the beacon messages have shorter range, i.e. the amount of vehicles that is able to receive beacons is affected by radio signal shadowing. Let us remember that the UV-CAST operation is based on information provided by the beacon messages, thus UV-CAST loses robustness. It is important to say that the low average packet delay (lower than 1s) of Slotted-1-Persistence is because it does not perform the tasks of store, carry and forward. Its performance is limited to direct relay through the dissemination of multi hop, which can be done very quickly.

As a next step, we evaluate the number of messages transmitted for the three radio shadowing models in Fig. 7. In the case free space (F), a large number of messages transmitted by UV-CAST (3 lines at the top of the Fig. 7), especially at high densities. This behavior is particularly expected, because UV-CAST uses beacon messages for information on traffic density. Also, because informed vehicles immediately forward the emergency message when they receive a beacon message from a neighbor node without the message. This results in redundant retransmissions, contention and many lost packets, which explains the high number of transmissions. Similar to the previous metrics, there is a difference between total messages transmitted in free space (F), shadows of buildings (B) and shadows of building and high vehicles (B+V). In those cases, the impact of the radio signal shadowing models provides opportunities due to a reduced channel load. This is evidenced by the decrease of the number of messages transmitted. The low number of messages transmitted (lower than 500) in Slotted-1-Persistence (3 lines at the bottom at Fig. 7) is because the protocol does not use beacons and because it has not the mechanism of opportunistic forwarding.

Finally, in Fig. 8 the average values for the Dissemination Efficiency (DE) are presented for the three radio shadowing models. As it can be seen in the case free space (F), DE values in the Slotted-1-Persistence protocol (3 lines at the bottom at Fig. 8) decrease in higher traffic density. Although

Slotted-1-Persistence was designed for high densities, the short propagation range and its high rate of packet losses do not allow to get a better performance. In contrast, as can be observed that in the case of UV-CAST (3 lines at the top of the Fig. 8), DE values have a tendency to grow at high density. Undoubtedly opportunistic forwarding mechanism allows to achieve a propagation distance greater and a higher delivery rate in comparison to Slotted-1-Persistence. Therefore, opportunistic forwarding performs better in the dissemination. This metric confirms that the dissemination is affected by the patterns of shadows (B and B+V). Consequently, we can see that Dissemination Efficiency is a metric that allows a clear view of the performance of a dissemination protocol.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have studied several proposals for dissemination of emergency messages in realistic urban scenarios. Several dissemination protocols have been presented in the literature, however due to frequent topology changes and dynamic traffic density, new protocols should be designed taking into account the special characteristics of realistic vehicular networks. We believe that dissemination protocols and general road safety applications should be evaluated considering scenarios where the shadows of buildings and tall vehicles complement the realism of the simulation.

As a future work, we plan to design a dissemination protocol that reduces the number of redundant messages, while ensuring a good coverage of spread and low latency in the message delivery. Furthermore, the proposed protocol will be evaluated together with other proposals of literature in realistic scenarios where human behavior is another variable to be considered.

ACKNOWLEDGMENT

This work was partly supported by the Spanish Government through projects TEC2013-47665-C4-1-R (EMergency Response In Smart COmmunities, EMRISCO), TEC2014-54435-C4-1-R (INcident monitoRING In Smart COmmunities. QoS and Privacy, INRISCO) and AGAUR Information Security Group (ISG) project - 2014 SGR 1504. Cristian Iza is recipient of a grant from Secretaria Nacional de Educación Superior, Ciencia y Tecnología SENESCYT. Ahmad Mezher is the recipient of a FI-AGAUR grant, from the Government of Catalonia.

REFERENCES

- [1] "Global status report on road safety 2013," <http://goo.gl/UIqA60>, accessed: 2015-03-22.
- [2] M. Picone, S. Busanelli, M. Amoretti, F. Zanichelli, and G. Ferrari, *Advanced Technologies for Intelligent Transportation Systems*, ser. Intelligent Systems Reference Library. Springer, 2015, vol. 139. [Online]. Available: <http://dx.doi.org/10.1007/978-3-319-10668-7>
- [3] S.-Y. Ni, Y.-C. Tseng, Y.-S. Chen, and J.-P. Sheu, "The broadcast storm problem in a mobile ad hoc network," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, ser. MobiCom '99. New York, NY, USA: ACM, 1999, pp. 151–162. [Online]. Available: <http://doi.acm.org/10.1145/313451.313525>
- [4] M. Boban, R. Meireles, J. Barros, P. Steenkiste, and O. Tonguz, "TV_r—tall vehicle relaying in vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1118–1131, 2014. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6529075>
- [5] C. Sommer and F. Dressler, *Vehicular Networking*. Cambridge University Press, November 2014.
- [6] "Veins, vehicular network simulations," <http://veins.car2x.org>, accessed: 2015-03-22.
- [7] "Omnet++, discrete event simulator," <http://www.omnetpp.org>, accessed: 2015-03-22.
- [8] "Sumo – simulation of urban mobility," <http://goo.gl/uvvD4N>, accessed: 2015-03-22.
- [9] "IEEE Guide for Wireless Access in Vehicular Environments (WAVE) - Architecture," IEEE.
- [10] "Openstreetmap," <http://www.openstreetmap.org/>, accessed: 2015-03-22.
- [11] C. Sommer, D. Eckhoff, R. German, and F. Dressler, "A computationally inexpensive empirical model of IEEE 802.11p radio shadowing in urban environments," in *Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on*, 2011, pp. 84–90. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5720204>
- [12] C. Sommer, S. Joerer, M. Segata, O. Tonguz, R. Lo Cigno, and F. Dressler, "How shadowing hurts vehicular communications and how dynamic beaconing can help," *IEEE Trans. Mobile Comput.*, to be published, early Access. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6920079>
- [13] M. Boban, T. Vinhoza, M. Ferreira, J. Barros, and O. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 15–28, 2011. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5678778>
- [14] "Propagation by diffraction," *ITU-R Rec. P.526-10*, February 2007.
- [15] S. Panichpapiboon and W. Pattara-Atikom, "A review of information dissemination protocols for vehicular ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 3, pp. 784–798, 2012. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5989903>
- [16] N. Wisitpongphan, O. Tonguz, J. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *IEEE Wireless Commun. Mag.*, vol. 14, no. 6, pp. 84–94, 2007. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4407231>
- [17] O. Tonguz, N. Wisitpongphan, and F. Bai, "Dv-cast: A distributed vehicular broadcast protocol for vehicular ad hoc networks," *IEEE Wireless Commun. Mag.*, vol. 17, no. 2, pp. 47–57, 2010. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5450660>
- [18] W. Viriyasitavat, O. Tonguz, and F. Bai, "Uv-cast: an urban vehicular broadcast protocol," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 116–124, 2011. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6069718>
- [19] R. S. Schwartz, H. Scholten, and P. J. M. Havinga, "A scalable data dissemination protocol for both highway and urban vehicular environments," *EURASIP J. Wireless Comm. and Networking*, vol. 2013, p. 257, 2013. [Online]. Available: <http://dx.doi.org/10.1186/1687-1499-2013-257>
- [20] J. A. Sanguesa, M. Fogue, P. Garrido, F. J. Martinez, J.-C. Cano, and C. T. Calafate, "Using topology and neighbor information to overcome adverse vehicle density conditions," *Transportation Research Part C: Emerging Technologies*, vol. 42, p. 113, May 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.trc.2014.02.010>
- [21] S. Zemouri, S. Djahel, and J. Murphy, "A fast, reliable and lightweight distributed dissemination protocol for safety messages in urban vehicular networks," *Ad Hoc Networks*, vol. 27, p. 2643, Apr 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.adhoc.2014.11.016>
- [22] K. Suriyapaiboonwattana, C. Pornavalai, and G. Chakraborty, "An adaptive alert message dissemination protocol for vanet to improve road safety," in *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*, 2009, pp. 1639–1644. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5277261>
- [23] K. N. Nakorn and K. Rojviboonchai, "Non-gps data dissemination for VANET," *IJDSN*, vol. 2014, 2014. [Online]. Available: <http://dx.doi.org/10.1155/2014/906084>

Certificado Digital Accesible para la e-Administración

Boni García¹, Ana Gómez¹, Rafael Conde¹, Yolanda Hernández², Miguel Ángel Valero¹

¹Universidad Politécnica de Madrid. Departamento de Ingeniería y Arquitecturas Telemáticas

Carretera de Valencia km 7, 28031 Madrid, España

boni@diatel.upm.es, agomez@diatel.upm.es, rafaelcondemelguizo@gmail.com, mavalero@diatel.upm.es

²ASPAYM Madrid. Camino de Valderribas 115, 28038 Madrid, España

terapiacupacional@aspaymmadrid.org

Resumen- La Administración electrónica (e-Administración) tiene como objetivo lograr que los servicios telemáticos públicos sean accedidos por los ciudadanos de forma cómoda y eficaz. Estos beneficios se acentúan para las personas con discapacidad física, ya que estos servicios permiten reducir tiempos de espera así como el número de desplazamientos. Sin embargo, su uso no está muy extendido en este colectivo, ya que a los posibles miedos propios de cualquier ciudadano hay que sumar las barreras propias de su discapacidad.

En este contexto, este grupo de investigación ha participado en el desarrollo de una solución que facilite a las personas con discapacidad el acceso a los servicios de la e-Administración. Este trabajo, desarrollado en cooperación con la Federación Nacional de Lesionados Medulares y Grandes Discapacitados Físicos, incluye la creación de un portal especialmente orientado a la gente con discapacidad, que les facilite la localización y acceso a los servicios ofrecidos por la e-Administración española. El uso de este portal se apoya en la autenticación digital de sus usuarios basada en certificados X.509, presentes en las tarjetas de identidad de las que disponen los ciudadanos españoles. Sin embargo, el análisis de uso realizado ha constatado que su uso supone una importante barrera de accesibilidad. En este artículo se propone una solución más accesible, basada en el empleo de un token criptográfico USB, capaz de ocultar a sus usuarios toda la problemática que conlleva el acceso a aplicaciones basadas en certificados, a la vez que les garantiza la seguridad requerida.

Palabras Clave: e-Administración, Accesibilidad, Certificado Digital, Token Criptográfico.

I. INTRODUCCIÓN

En el año 2006 la Unión Europea lanzó su i2010 e-Government Action Plan [1], posteriormente revisado en The European eGovernment Action Plan 2011-2015 [2], destinado a modernizar y hacer más eficaces los servicios públicos de los estados miembros de la UE para reducir la carga de burocracia e ineficacia a la que se enfrentaban los ciudadanos. Como consecuencia, todos los estados de la UE están embarcados en la generación de aplicaciones de gobierno electrónico, enfocadas por un lado, a responder a las necesidades y mejora de su propia gestión interna y, por otro, a atender las demandas de la sociedad con respecto a transparencia, información y prestación de servicios.

El desarrollo de nuevos servicios de Administración electrónica conlleva la necesidad una identificación digital que permita autenticarse inequívocamente al realizar las gestiones a través de Internet. Para atender esta exigencia, muchos países europeos están promoviendo entre sus ciudadanos el uso de certificados X.509 para sus relaciones

con la Administración, tanto en su versión software como a través de dispositivos portables seguros. En España, al igual que en otros países europeos, se ha optado por añadir una funcionalidad adicional a la tarjeta de identidad, ya existente y obligatoria para todos los ciudadanos, mediante la incorporación de un chip. Este chip contiene los certificados X.509 y las claves necesarios para llevar a cabo las transacciones con la e-Administración de forma segura.

Las Administraciones públicas españolas permiten la interacción bidireccional y la tramitación completa de los servicios públicos con niveles altos, superiores a la media de la UE27 [3]. Sin embargo, a pesar de que en la actualidad el 80% de la población española mayor de 14 años, dispone de tarjeta de identidad digital, solo un número reducido de ciudadanos utilizan los servicios que pone a disposición la Administración Pública. Las razones hay que buscarlas, por una parte, en la desconfianza de los ciudadanos ante el uso de nueva herramienta y, por otra, y quizás más importante, por la dificultad de uso.

En este escenario, las personas con discapacidad, alrededor del 8,5% de la población española, se enfrentan a una doble dificultad para acceder a los servicios de la Administración Pública, ya que por un lado encuentran barreras para acceder de forma presencial a los edificios o entidades y, por otro, una dificultad todavía mayor para acceder vía telemática. Para evitar esta situación es preciso emprender acciones que minimicen la brecha digital y maximicen la accesibilidad de estas personas a los servicios telemáticos, que tanto pueden contribuir a mejorar su calidad de vida.

Esta pieza de investigación propone una solución para facilitar a las personas con discapacidad el acceso a los servicios de e-Administración. Para ello proponemos el uso token criptográfico USB, capaz de ocultar a sus usuarios toda la problemática que conlleva el acceso a aplicaciones basadas en certificados. El trabajo ha sido desarrollado en cooperación con la Federación Nacional Aspaym (Asociación de Lesionados Medulares y Grandes Discapacitados Físicos) e incluye el desarrollo de un portal, especialmente orientado a la gente con discapacidad, que les facilite la localización y acceso a los servicios ofrecidos por la Administración española.

El resto del artículo está estructurado como sigue. En la sección 2 se presenta el estado de la técnica en relación con

la e-Administración y el uso del certificado digital. La sección 3 recoge el análisis cualitativo realizado para detectar cuáles son las principales dificultades que tienen las personas con discapacidad para utilizar los servicios de la e-Administración. Como consecuencia de este análisis, la sección 4 analiza las diferentes opciones barajadas para solucionar los problemas detectados. La sección 5 presenta la solución finalmente adoptada. La sección 6 describe las pruebas realizadas para verificar la viabilidad de la propuesta. Por último, la sección 7 presenta las conclusiones y trabajo futuro.

II. ESTADO DEL ARTE

Esta sección proporciona un análisis de situación de los certificados digitales en los que está enmarcado este trabajo de investigación. Además, se estudian trabajos relacionados disponibles en la literatura.

A. Certificados Digitales

Un certificado de clave pública, por lo general conocido simplemente como “certificado”, es una declaración firmada digitalmente que enlaza el valor de una clave pública con la identidad de una persona, un dispositivo o un servicio que posee la clave privada correspondiente [4].

Una tarjeta de identidad electrónica (eID, *Electronic Identity Card*) es un documento emitido por una autoridad oficial para permitir la identificación de usuarios de forma online y offline. Los países que actualmente emiten los eID son Bélgica, Alemania, Italia, Países Bajos, Pakistán, Rumanía, Estonia y España. En el caso de España, el eID es conocido como DNIE (Documento Nacional de Identificación Electrónico). El Gobierno español ofrece el DNIE con el fin de proporcionar autenticación y capacidad de firma electrónica. El DNIE sigue la norma ISO 7816, que a su vez es una evolución del estándar PKCS#15 [5].

En enero de 2015, España ha emitido más de 38 millones de DNI electrónicos¹. El DNIE contiene dos certificados X.509 de ciudadano, uno de autenticación y otro de firma, así como las claves privadas asociadas a cada uno. Según la encuesta sobre equipamiento y uso de tecnologías de la información y comunicación del Instituto Nacional de Estadística de España, sólo un 4,7% de los españoles hacen uso del DNIE en sus relaciones con la Administración electrónica [6]. Algunos factores a tener en cuenta sobre estas escasas cifras de uso son los siguientes:

- No todas las personas que tienen un DNIE con el certificado generado han guardado la clave de utilización.
- No todos los ciudadanos disponen de un ordenador configurado (drivers y lector de tarjetas) para poder usar el DNIE.
- Los certificados del DNIE caducan cada 20 meses y requieren renovación presencial en las oficinas de la Policía.

En vista de los inconvenientes que el uso del DNIE presenta para los trámites con la Administración electrónica, es posible utilizar otro tipo de certificado digital de usuario. Entre la veintena de certificados electrónicos vigentes en la actualidad en España, el más popular es el certificado Ceres

emitido por la Fábrica Nacional de Moneda y Timbre (FNMT) de clase 2CA.

El uso de un certificado electrónico brinda la posibilidad de realizar multitud de trámites y gestiones on-line, ahorrando tiempo de desplazamientos y llamadas. Esto confiere un valor indudable a personas con movilidad reducida y sus familiares o cuidadores. Por una parte, existe la posibilidad de realizar los trámites cualquier día de la semana y a cualquier hora, sin necesidad de desplazarse o involucrar a una tercera persona, con el consiguiente ahorro de tiempo (debe tenerse en cuenta que el tiempo que necesita una persona con movilidad reducida aumenta con respecto al tiempo normal de cualquier ciudadano).

Aunque existen numerosas oportunidades de tramitar servicios de forma telemática, hasta ahora no son mayoritariamente usados por las personas con discapacidad. Las razones hay que buscarlas en la dificultad que entraña su uso para una persona no familiarizada con el manejo de los certificados electrónicos, base en la cual se apoyan todos los servicios ofrecidos por la Administración. Hay que aclarar que esta dificultad no sólo atañe a las personas con discapacidad, sino que afecta a la mayoría de los ciudadanos. Sin embargo, resulta más significativa en un colectivo que podría resultar enormemente beneficiado con su uso.

B. Trabajos relacionados

En el estudio de la literatura previo a la ejecución investigación, no se han encontrado trabajos similares al que presenta este artículo. Este hecho en sí mismo ya es una buena razón para ser llevado a cabo. A continuación se enumeran brevemente algunas de las referencias relacionadas encontradas.

Heichlinger y Gallego describen la introducción de la tarjeta de identidad electrónica en España (DNIE) [7]. En las conclusiones de este artículo se puede constatar que el caso del DNIE en España ha supuesto un desarrollo bien diseñado e implementado, pero con una asimilación bastante lenta, así como un uso muy bajo de la función de identificación electrónica para la autenticación en línea. Siguiendo con el DNIE, Ruiz-Agundez y Bringas proponen un servicio de autenticación mediante tarjetas de identidad electrónica para VoIP. Los autores proponen como líneas de trabajo futuro el uso de otros eIDs a parte del DNIE [8].

El artículo de Olsen et al [9] realiza un estudio de las ventajas e inconvenientes de la Administración electrónica frente a la tradicional. Los resultados de esta investigación sugieren que después de la inversión inicial y la curva de aprendizaje, la Administración electrónica supone un importante ahorro de tiempo y esfuerzo. Las desventajas principales son el costo inicial del software y el apoyo técnico necesario para configurar y operar los programas con éxito.

En el libro “Constructing Accessible Web Sites”, los autores explican diferentes técnicas para la construcción de aplicaciones web accesibles, desde el punto de vista de los contenidos, navegación, entrada de datos, pruebas, herramientas, etc. [10] Pero en esta referencia no se detiene a estudiar la seguridad desde el punto de vista de la accesibilidad, dejando de manifiesto que existe un vacío de conocimiento en esta mezcla interdisciplinar (accesibilidad más seguridad).

¹ <http://www.dnielectronico.es/>

Por tanto, será preciso buscar una solución que combine las ventajas del certificado digital y la accesibilidad. En el siguiente apartado se indica la metodología seguida para realizar un estudio preliminar que permitirá detectar los aspectos que influyen negativamente en el acceso a la e-Administración usando aplicaciones web seguras.

III. METODOLOGÍA

El primer paso para lograr que las personas con discapacidad hagan uso de la Administración electrónica ha consistido en averiguar cuáles son las dificultades específicas a las que deben enfrentarse. Para ello se diseñó un análisis cuantitativo para el que fue preciso seleccionar una muestra significativa de los servicios públicos ofrecidos por la Administración, elaborar una muestra de usuarios representativa con la población a quien va dirigida la solución tecnológica y diseñar el trabajo de campo para que los usuarios probasen los servicios telemáticos. A continuación se comenta cómo se llevó a cabo esta encuesta y los resultados obtenidos.

A. Selección de servicios

Para seleccionar los servicios se llevó a cabo un trabajo de documentación y análisis deliberativo con personas con discapacidad. En primer lugar, se consultó la principal documentación oficial sobre evaluación internacional de Administración electrónica. A nivel global se consultó la United Global Nations Global E-Government Survey [11], a nivel europeo el Indicador sobre Administración Electrónica de la Oficina Estadística de las Unión Europea Eurostat [12] y a nivel español la Ley de Acceso electrónico de los Ciudadanos a los Servicios Públicos [13] que define los derechos que los servicios de e-Administración deben permitir.

A partir de la información recopilada se creó un cuestionario donde se recogían los principales servicios identificados para discriminar por uso, más o menos habitual. Finalmente, tanto las conclusiones de la documentación como los resultados del cuestionario se debatieron en tres grupos de discusión formados por personas con cada una de las discapacidades incluidas en la investigación: daño cerebral, lesión medular y parálisis cerebral. El resultado final fue la selección de 10 servicios básicos para hacer el estudio, considerados muy útiles para las personas con discapacidad, pertenecientes a todos los niveles de Administración del estado.

B. Perfil de usuarios

Los tres perfiles que se han tenido en cuenta para las pruebas fueron parálisis cerebral (PC), daño cerebral (DC) y lesión medular (LM). El denominador común de estos tres perfiles es la dificultad de movilidad asociada a la situación de discapacidad. Es por ello que los servicios telemáticos accesibles proporcionan una solución a un problema real que se encuentran las personas con discapacidad física, al mejorar los tiempos de espera en las gestiones, reducir el número de desplazamientos y facilitar la comunicación con la administración.

C. Muestra de usuarios

Se elaboró una muestra de usuarios representativa para probar los servicios seleccionados. La intención de la muestra

no fue obtener datos aplicables al resto de la población, por lo que no se trató de una muestra aleatoria, sino de una muestra estructural, entendiéndose por este término una muestra con tendencia a la saturación estructural y no a la representación estadística [14]. El objetivo de la muestra fue reunir todos los perfiles que se consideran relevantes a la hora de identificar barreras en el acceso a los servicios de administración pública electrónica.

La primera variable estructural ha sido el nivel de dependencia. Debido a la homogeneidad de las discapacidades entre los colectivos que participan en la investigación, se ha decidido adoptar la escala de la Ley de Dependencia española [15]. Esta Ley distingue entre tres niveles de dependencia en función de las capacidades del individuo para ser autónomo en las actividades de la vida diaria: dependencia moderada, dependencia severa y gran dependencia. Al tratarse de una tipología elaborada con el objetivo de evaluar todo tipo de discapacidades, esta herramienta ha permitido estructurar la muestra con una variable objetiva transversal a las diferentes discapacidades.

En segundo lugar, se han definido tres niveles de experiencia en el uso de tecnología, previa a la participación en la investigación y se han descartado los individuos sin experiencia previa. El resto ha sido catalogado en tres niveles: bajo (utiliza las tecnologías de forma esporádica), medio (utiliza las tecnologías habitualmente como usuario) y alto (utiliza las tecnologías habitualmente y tiene conocimientos avanzados como instalar un programa o solucionar un problema del sistema operativo).

Por último, se ha buscado la paridad entre mujeres y hombres.

Con estos tres criterios estructurales, se ha establecido un mínimo de dos individuos por perfil. Al haber dos individuos por perfil, obtenemos una muestra estructural de 36 personas, estructurada de la siguiente manera:

- 12 personas por cada grado de dependencia.
- 12 personas por cada nivel de experiencia previa en el uso de las nuevas tecnologías.
- 24 personas de cada género.
- Cada una de las personas ha realizado dos pruebas, por lo que en total se han realizado 72 pilotajes. De este modo, cada servicio ha sido probado un mínimo de 7 veces.
- La mitad de las pruebas se realizaron con identidad digital y la otra mitad sin ella.

Esta composición de la muestra estructural permite considerar que se ha cumplido el criterio de saturación cualitativa para la validez de la investigación [14].

D. Trabajo de campo

Este trabajo se llevó a cabo en colaboración con la Federación Nacional de Lesionados Medulares y Grandes Discapacitados Físicos (ASPAYM). Los pilotajes tuvieron lugar en las sedes colaboradoras españolas de Madrid, Toledo, Cuenca, Albacete y Murcia, y en los hogares de algunos usuarios. En cada pilotaje se le pedía al usuario que realizase una gestión con cada una de las dos web que le tocaba evaluar. La persona observadora no intervenía si el usuario no precisaba de ayuda. La información del pilotaje se ha recogido de tres formas:

- Observación directa de los responsables de los pilotos.
- Un cuestionario sobre la experiencia de accesibilidad al final de cada pilotaje. El cuestionario se ha denominado

ACDM-16 y se ha elaborado basándose en las pautas de accesibilidad WCAG 2.0.

- Una entrevista no estructurada realizada tras el cuestionario sobre las cuestiones observadas durante la prueba.

E. Análisis de resultados y conclusiones

Tras el trabajo de campo, se ha procedido al análisis de la información recogida. En primer lugar, se ha procedido a un análisis descriptivo de los resultados del cuestionario ACDM-16 mediante cruces bivariantes entre la experiencia de usabilidad y las variables estructurales de la muestra: nivel de dependencia, experiencia de uso de las nuevas tecnologías y género. Los resultados del análisis descriptivo se han analizado en base a la información obtenida en la observación y en las entrevistas informales.

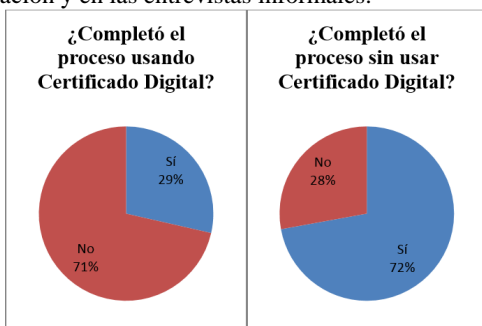


Fig. 1. Resultados relacionados con el uso de certificados.

Se ha comprobado que la variable que más influye en la experiencia de accesibilidad a los servicios telemáticos de la administración ha sido el tipo de discapacidad, fundamentalmente a través de las características de desarrollo cognitivo asociadas a cada una de las discapacidades. Estos resultados han llevado a la conclusión de que la solución tecnológica debe incluir elementos relacionados con la accesibilidad cognitiva (lectura fácil, minimizar las rutas de enlaces, pictogramas, etc.) junto a los elementos de accesibilidad técnica relacionados con el manejo del certificado digital o la experiencia de navegación.

Otro de los resultados más destacables de este estudio ha sido la constatación de que el uso de los certificados presenta grandes dificultades para las personas con discapacidad, tanto si está instalado en el navegador como si está disponible a través de la tarjeta de identidad del ciudadano (ver figura 1).

En primer lugar, se detectan dificultades para el uso inicial de la tarjeta de identidad, que aunque no específicas de este colectivo, sí contribuyen a que las personas que intenten usar los servicios telemáticos ofrecidos por la Administración desistan de ello. Esto es debido a que antes de la puesta en marcha es necesario disponer de un lector de tarjetas (para el que a veces hay que instalar un driver específico) y descargar el módulo criptográfico en la página oficial del organismo que expide el documento de identidad. También conviene cambiar el PIN que viene preinstalado en la tarjeta de identidad. Este PIN, formado por una serie de 16 caracteres alfanuméricos aleatorios, resulta inmanejable, por lo que se recomienda cambiarlo a uno que tenga entre 8 y 16 caracteres, que incluya letras mayúsculas y minúsculas y caracteres.

La primera dificultad en su uso, viene motivada por las reiteradas peticiones al usuario para que teclee su número de

PIN cuando hace un trámite con la Administración (tantas como accesos a la tarjeta requiera la aplicación). Sin embargo, el principal obstáculo para un usuario no familiarizado con el manejo de certificados, viene ocasionado por las alertas de seguridad del navegador, en forma de ventanas emergentes e intimidatorias (ver figura 2), que avisan de la falta de confianza del navegador en el servidor web al que se está accediendo. Para evitar estas alertas, es preciso instalar los certificados raíz de las Autoridades de Certificación que han expedido los certificados que intervienen en la transacción (el nuestro y el del servidor). Las sedes electrónicas suelen disponer de ayudas para que los usuarios atiendan estos mensajes, pero sus explicaciones deben adaptarse al sistema operativo, navegador y versión que éstos utilicen. Dadas las dificultades que encuentran los usuarios, es frecuente que una vez que se ha intentado infructuosamente realizar una transacción, abandonen definitivamente la idea de usar el certificado digital.

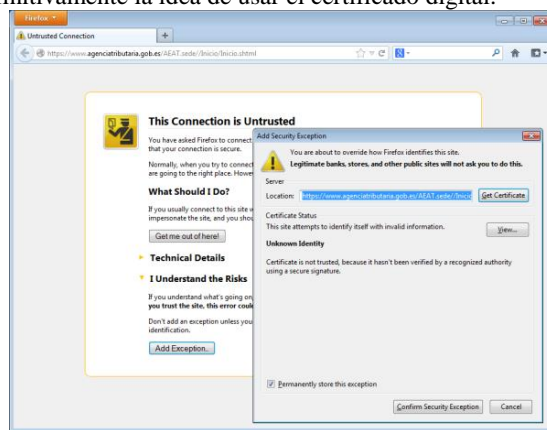


Fig. 2. Ejemplo de Conexión no Verificada en Firefox

A partir de las dificultades detectadas se ha decidido trabajar en dos campos complementarios. Por una parte, el diseño de un portal web accesible, específicamente orientado a personas con discapacidad que les permita acceder de forma cómoda a los servicios que más demanda este colectivo. En segundo lugar, diseñar una solución que facilite las interacciones con la Administración mediante el empleo de un dispositivo alternativo al documento de identidad digital, que sea más cómodo y usable para los usuarios. Esta última solución es la que se recoge en este artículo.

IV. CERTIFICADO DIGITAL ACCESIBLE

El dispositivo accesible que proponemos como alternativa al DNIe para almacenar la identidad digital para los usuarios se ha bautizado con el nombre de CDA² (Certificado Digital Accesible). Los requisitos de este sistema son los siguientes:

1. Se usará un token criptográfico para almacenar el certificado digital del usuario.
2. El token además tendría un navegador web portable completamente configurado y listo para su uso en aplicaciones web seguras.

² <http://bonigarcia.github.io/cda/>

3. Cuando se conectase el token por USB al ordenador del usuario, se debería lanzar automáticamente dicho navegador portable.

4. Cuando el usuario cerrase el navegador, se expulsaría automáticamente de forma segura el dispositivo USB.

A. Sistema Operativo

CDA será ejecutado en ordenadores con sistema operativo de escritorio. Según las estadísticas de la compañía Net Applications, Microsoft Windows es el sistema operativo predominante en este mercado [16] (ver figura 3). Por esta razón, CDA será implementado para ser compatible con sistemas Microsoft Windows.

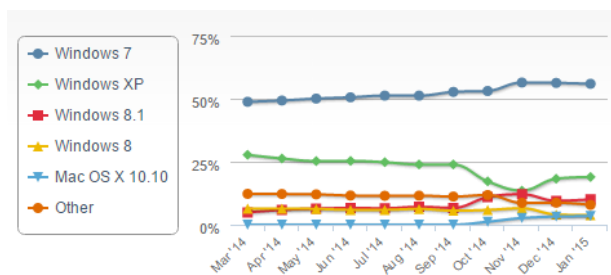


Fig. 3. Cuota de uso de Sistemas Operativos de Escritorio

B. Navegador

Para garantizar la facilidad de uso en la navegación de aplicaciones web seguras, hay que tener muy en cuenta los certificados que maneja el navegador. Un navegador puede usar dos tipos de certificados:

1. Certificado digital de usuario. Es un documento firmado electrónicamente que confirma la identidad del usuario. Como se ha estudiado en la sección 2, en España, actualmente los certificados electrónicos emitidos por entidades públicas son el DNIe y el certificado Ceres.

2. Certificados digitales de servidores o autoridades de confianza. Emitidos por entidades de certificación intermedias y entidades de certificación, requeridos para la correcta verificación del certificado.

El certificado digital de usuario será incorporado al token criptográfico. Los certificados digitales de servidores deberán ser actualizados y gestionados de forma transparente al usuario por el sistema CDA. La forma en que los diferentes navegadores actuales gestionan este tipo de certificados varía en función de cada uno.

Los navegadores más usados según las estadísticas del W3C son Internet Explorer, Firefox, Chrome, Safari, y Opera (ver figura 4) [17]. De estos navegadores, las versiones para Windows de Internet Explorer, Chrome y Safari usan el almacén de claves nativo de Windows. Por ese hecho estos tres navegadores no pueden formar parte de CDA, ya que el almacén de claves debe ser portable junto al navegador en el token. Las alternativas posibles son Firefox y Opera, que usan un almacén de certificados propio y por tanto portable.

Firefox usa el almacén de certificados de NSS³ (Network Security Services). NSS es un conjunto de bibliotecas diseñadas para permitir el desarrollo multiplataforma de

aplicaciones distribuidas y seguras. La gran ventaja de NSS es que ofrece una gran variedad de herramientas de desarrollo, muy útiles a la hora de facilitar la implementación. Por esta razón, se consideró NSS como mejor opción frente al almacén propio de Opera, y por lo tanto el navegador elegido para el sistema CDA fue Firefox.

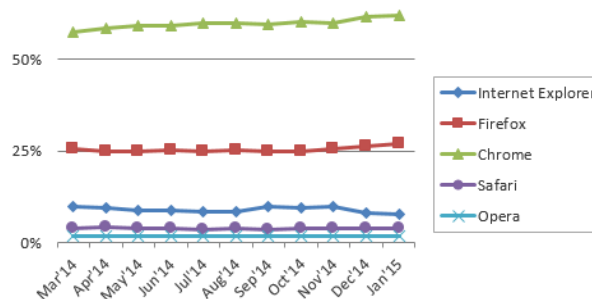


Fig. 4. Cuota de uso de Navegadores web

Ahondando más en los detalles de NSS, el almacén de certificados de Firefox es almacenado en una base de datos de tipo Berkeley DB instrumentado por el fichero cert8.db [18]. Para garantizar la facilidad de uso del navegador Firefox de CDA, la actualización de esta base de datos será realizada automáticamente. Es por ello que se añadió un servidor web al sistema CDA en el cual se aloja una base de datos cert8.db actualizada. El token se conecta de forma segura al servidor para realizar la actualización de esta base de datos de manera transparente al usuario. De este modo se consigue evitar situaciones en las que el navegador no puede confirmar que la conexión con determinadas aplicaciones no sea segura (ver figura 2).

C. Token Criptográfico

Un token criptográfico es un dispositivo electrónico que proporciona servicios de autenticación y autorización. El token que se usará para implementar el CDA debe ser capaz de almacenar el certificado digital de usuario de manera segura.

Existen multitud de tokens criptográficos disponibles en el mercado. En base a los requisitos definidos para CDA, el dispositivo elegido fue el token criptográfico iAM desarrollado por la empresa Bit4id⁴. Se trata de un dispositivo USB basado en un microprocesador criptográfico que incorpora una memoria micro-SD con aplicaciones para firma electrónica, cifrado, y verificación de documentos firmados electrónicamente. El dispositivo incluye una tarjeta SIM en la que pueden ser almacenados certificados personales de usuario. El acceso a este certificado está protegido mediante códigos PIN/PUK.

D. Escenario de uso

Para llevar a cabo los requisitos 2 y 3 propuestos al principio de esta sección, la solución más sencilla era la creación de una aplicación que se ejecute al conectar el token por USB en la máquina del usuario. Dado que la solución se desarrollaba para sistemas Windows, se podría haber usado el componente Autorun de Windows. Desde Windows 95,

³ <http://www.mozilla.org/projects/security/pki/nss/>

⁴ <http://www.bit4id.com/en/>

Autorun permite la ejecución de determinadas acciones al insertar un medio extraíble. Las acciones llevadas a cabo usando este método son listadas en un fichero llamado autorun.inf contenido en el medio conectado [19]. Muy ligado al Autorun, desde Windows 98 se introdujo el Autoplay. Esta característica de Windows examina el contenido de un dispositivo extraíble al ser conectado en el sistema y ofrece una serie de opciones para su ejecución [20].

Frecuentemente, el Autorun de Microsoft ha sido usado para transmitir malware. Por esta razón, el Autorun ha sido deshabilitado para dispositivos USB desde Windows 7. Un estudio de Microsoft de 2011 encontró que el 26% de infecciones en sistemas Windows provenía de la instalación de malware vía dispositivos USB [21]. Debido a esto, en febrero de 2011 se hizo pública la actualización de seguridad KB971029⁵ para sistemas Windows XP y posteriores, mediante la cual el Autorun para USB quedaba deshabilitado. Por todo lo expuesto y en base a los requisitos establecidos, se propuso un diseño del sistema CDA usando los siguientes componentes:

- CDA-Listener. Componente software que se instala en la máquina del usuario. Tiene programado un evento de escucha para detectar la conexión por USB del token en el sistema. Cuando detecta dicha conexión lanza el componente CDA-Autorun que está almacenado en el token. Además, contiene un componente que detecta la finalización del navegador portable del token. En ese momento pone en funcionamiento la expulsión del token USB de forma segura.

- CDA-Autorun. Componente software que se almacena en el token. Realiza la actualización de la base de datos de certificados del Firefox portable del token. Para ello se conecta de forma segura a un servidor web que contiene una base de datos de certificados actualizada. Después de realizar esta actualización, lanza el Firefox para que el usuario pueda usar este navegador usando su certificado digital.

El procedimiento de uso del token CDA se ilustra en la figura 5. Dado que el objetivo principal de este trabajo es la accesibilidad, la pieza fundamental del escenario es el usuario. Los pasos que el usuario tendrá que realizar para usar el token son los siguientes:

1. Instalación del componente CDA-Listener en el sistema operativo del usuario (paso etiquetado como "1" en el diagrama). El instalador de este componente se ha implementado mediante un sencillo Wizard que requerirá la atención del usuario simplemente para aceptar la instalación a través de varias pantallas. Este paso sólo será realizado por el usuario una vez, ya que CDA-Listener quedará instalado en el sistema creando una entrada en el registro de Windows para que el programa se arranque al iniciar sesión.

2. Conexión del token por USB. En este instante, y dado que CDA-Listener está ya instalado en el sistema, este componente detecta que el token ha sido conectado, y lanza la ejecución del componente CDA-Autorun del token (paso "2a"). CDA-Autorun se conecta a un servidor web para actualizar la base de datos de certificados (paso "2b"). Para la conexión segura utiliza una serie de certificados de cliente X.509. Como respuesta obtendrá la base de datos de

certificados Firefox actualizado (fichero cert8.db). Las excepciones de seguridad son almacenadas por Firefox en un fichero de texto llamado cert_override.txt. Ambos ficheros son automáticamente actualizados (paso "2b-r"). En este momento es cuando se lanza el Firefox portable (paso "2c").

3. El usuario utiliza el navegador de forma segura y con su certificado de usuario. Una vez que decida finalizar la navegación, cerrará el navegador Firefox. En ese instante el componente "Firefox Listener" de CDA-Listener llamará al componente "USB Ejector" para expulsar el dispositivo USB de forma segura (paso "3a"). Por último se informará al usuario que puede desconectar el token (paso "3b").

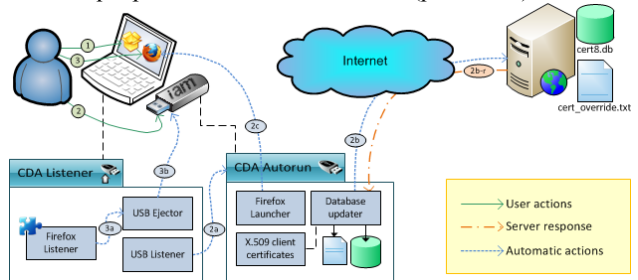


Fig. 5. Escenario de uso

V. IMPLEMENTACIÓN

Para la implementación de CDA-Listener y CDA-Autorun se eligió el lenguaje de programación C#. Esta decisión viene determinada por la plataforma en la que se va a ejecutar el sistema CDA, esto es, Windows. C#, al ser un lenguaje de desarrollo hecho para y por Microsoft, ofrece un acceso avanzado a todas las funciones de Windows. Este hecho será especialmente útil a la hora de desarrollar la escucha del dispositivo USB.

A. CDA-Listener

Este componente ha sido desarrollado en C# como una aplicación que estará visible a través de un icono de la barra de tareas de Windows. Está implementado como una aplicación de instancia única, lo cual significa que sólo estará presente una vez en el sistema.

1. USB Listener

La detección de la conexión de dispositivos USB (componente "USB Listener" en la figura 5) se realiza mediante un evento WQL (WMI Query Language) tal y como se describe a continuación.

Listado I

PROGRAMACIÓN DE LA ESCUCHA DE EVENTOS USB

```
stringusbScope = @"root\CIMV2";
stringusbQueryString = "SELECT * FROM
_InstanceOperationEvent WITHIN 3 WHERE
TargetInstance ISA 'Win32_DiskDrive'";
usb_listener = new
ManagementEventWatcher(usbScope,
usbQueryString);
usb_listener.EventArrived += new
EventArrivedEventHandler(usb_connected);
usb_listener.Start();
```

2. Firefox Listener

En una primera implementación de CDA-Listener, esta pieza se desarrolló mediante otra consulta WQL que escuchaba la finalización del proceso del Firefox portable del

⁵ <http://support.microsoft.com/kb/971029>

token. Si bien esta solución funciona correctamente, podría plantear cierto problema de uso de cara al usuario, ya que desde el momento en el que el usuario cierra el navegador hasta el momento en el que el sistema termina el proceso, transcurren un lapso de tiempo en el que el usuario no sabe que el dispositivo está siendo expulsado.

Para mejorar este funcionamiento, se ha implementado esta pieza como un complemento del Firefox portable del token. De esta forma es el propio navegador el que avisa que el usuario lo ha cerrado. En ese momento hará una llamada al componente "USB Ejector" de CDA-Listener para que inicie la desconexión segura del dispositivo USB. Entre tanto, un splash informará al usuario de la operación de expulsado seguro del token en curso.

Para la implementación de esta pieza se ha usado la herramienta online de desarrollo de complementos de Mozilla, el Add-on Builder⁶.

3. USB Ejector

Para la ejecución segura del dispositivo USB se han incorporado al componente CDA-Listener la herramienta de código abierto USB Disk Ejector⁷. Esta herramienta es portable, flexible, y puede ser llamada con diferentes argumentos desde la línea de comandos. Pero la razón principal de la elección de esta herramienta es el hecho de que antes de expulsar de forma segura el dispositivo USB, detecta las aplicaciones que han sido lanzadas desde dicho dispositivo y fuerza su finalización.

B. CDA-Autorun

El componente CDA-Autorun es responsable de la actualización de la base de datos de certificados y lanzamiento del Firefox portable.

Para garantizar una comunicación segura entre el token CDA y el servidor web que almacena los ficheros cert8.db y cert_override.txt usaremos certificados X.509. Estos certificados han sido generados con la herramienta de código abierto OpenSSL⁸.

El sistema CDA requiere el mantenimiento actualizado de la base de datos de certificados en un servidor web (cert8.db y cert_override.txt). Para la descarga de estos ficheros se emplea la herramienta de código abierto cURL⁹ con soporte SSL. La sintaxis del comando para descargar la base de datos SSL del servidor web donde se ha copiado la base de datos maestra cert8.db está descrito en el listado 2.

Listado II

COMANDO PARA DESCARGAR DE FORMA SEGURA LA BASE DE DATOS DE CERTIFICADOS

```
curl --remote-time --remote-name --ciphers RSA
--cert cda-client.crt --key cda-client.key --
cacertcda-ca.crt
https://innova.diatel.upm.es/db.ssl/cert8.db
```

⁶ <https://builder.addons.mozilla.org/>

⁷ <http://quickandeasysoftware.net/software/usb-disk-ejector>

⁸ <http://www.openssl.org/>

⁹ <http://curl.haxx.se/>

La base de datos descargada debe ser fusionada con la base de datos local del token. De esta forma se instalan los certificados que no existan de forma local y se actualizan los existentes. En el caso del fichero de texto cert_override.txt, esta fusión de contenidos se realiza comparando línea a línea el fichero descargado del servidor con el local del token. En el caso de cert8.db, se usa la herramienta certutil del paquete de herramienta de base de datos de certificados (Certificate Database Tool) de NSS.

VI. VERIFICACIÓN

Una vez desarrollados todos los componentes propuestos para CDA, se ha configurado el token adecuadamente y con el servidor listo, se han realizado una serie de experimentos de laboratorio para verificar la viabilidad funcional de la propuesta. Para ello se han usado diferentes máquinas virtuales para los diferentes sistemas operativos Windows (XP, Vista, 7, 8 y 8.1).

Estas máquinas virtuales se instalaron en un sistema de virtualización VMWare Workstation 8.0. Se usó el sistema de snapshots de VMWare para almacenar los diferentes estados de prueba a partir de los cuales se han evaluado las diferentes máquinas virtuales.

Para cada una de esas 5 máquinas virtuales Windows, se realizaron pruebas con diferentes antivirus instalados. Según la compañía de estadísticas de Internet Opswat, los antivirus más utilizados enero de 2015 han sido Microsoft Security Essentials y Avast Free Antivirus [22] (ver figura 6). Por esta razón se realizaron pruebas para Windows con cada uno de estos antivirus, más otro escenario correspondiente a un sistema Windows sin antivirus.

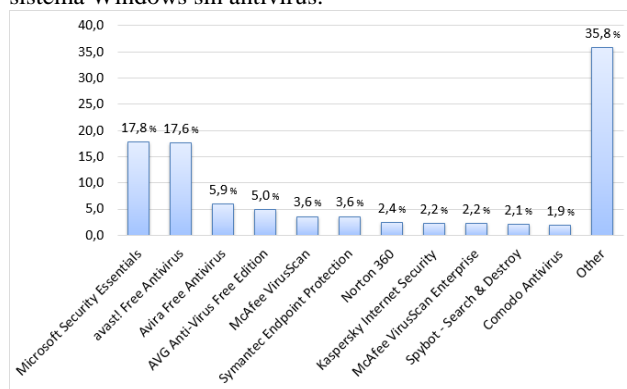


Fig. 6. Cuota de uso de Antivirus

De esta manera, se dispuso de 13 snapshots en los que realizar las pruebas. Esta cifra es el resultado de multiplicar el número de configuraciones de antivirus por el número de sistemas operativos y restar un escenario, debido a que no es posible probar Windows 8 y 8.1 sin antivirus, ya que viene por defecto con la nueva versión de Security Essentials, llamada Windows Defender. Para cada uno de estas 13 snapshots, el procedimiento de prueba fue el siguiente:

1. Instalar CDA-Listener.
2. Conectar token iAM.
3. Comprobar que se ejecuta automáticamente CDA-Autorun y se actualizan los certificados del navegador Firefox portable.

4. Comprobar que el Firefox portable no entra en conflicto con un Firefox previamente instalado en el sistema.
5. Navegar con el Firefox portable usando el certificado digital de usuario del token.
6. Cerrar el Firefox portable.
7. Comprobar que se expulsa automáticamente.

La tabla 2 recoge los resultados de las pruebas. El símbolo “✓” de esta tabla significa que todas las pruebas realizadas fueron satisfactorias. Por otro lado, el símbolo “X” significa que ocurrieron situaciones problemáticas en las pruebas.

Tabla II
RESULTADOS DE LAS PRUEBAS

	Sin antivirus	Security Essentials/ Windows Defender	Avast
Windows XP	✓	✓	✓
Windows Vista	✓	✓	X
Windows 7	✓	✓	X
Windows 7		✓	X
Windows 8.1		✓	X

Los problemas detectados tienen que ver con el uso del antivirus Avast. La razón de estos problemas es que Avast por defecto no confía en los componentes que componen el sistema, como CDA-Listener y CDA-Autorun. Por esta el usuario tiene que decirle expresamente que estos componentes son seguros, y de este modo continúa la ejecución de los mismos concluyendo correctamente. La solución a este problema consiste en excluir los componentes de CDA del análisis de Sandbox de Avast.

VII. CONCLUSIONES Y TRABAJO FUTURO

Las TIC representan una gran oportunidad para la capacitación en el acceso universal a la información y pueden convertirse en un importante aliado de las personas dependientes y sus familias. Sin embargo, las barreras existentes para el acceso a los servicios telemáticos que afectan a la población general, muchas veces se convierten en insalvables para colectivos de gente con discapacidad. Para evitar esta situación es preciso emprender acciones que minimicen la brecha digital y maximicen la accesibilidad de estas personas a los servicios telemáticos, que tanto pueden contribuir a mejorar su calidad de vida.

En esta línea se enmarca la solución recogida en este artículo, mediante el desarrollo del token criptográfico CDA, que permite un seguro y cómodo acceso a los servicios ofrecidos por la Administración, y que resulta especialmente útil para las personas con discapacidad. Este token oculta a los ciudadanos toda la problemática que conlleva el uso de certificados, de manera que proporciona una solución accesible para que las personas con discapacidad puedan beneficiarse de los servicios ofrecidos por la Administración electrónica.

Si bien se ha desarrollado un trabajo de verificación a nivel técnico de lo propuesta, todavía está pendiente un trabajo de validación entre el colectivo de personas con discapacidad de ASPAYM. La puesta en marcha del nuevo portal accesible para acceder a los servicios ofrecidos por la Administración, junto con el uso del token criptográfico facilitará a este colectivo la realización de los trámites

administrativos cómodamente y permitirá verificar en qué medida el empleo de CDA cumple con las expectativas previstas.

AGRADECIMIENTO

Este artículo ha sido desarrollado en el ámbito del proyecto ASTIC (Accesibilidad en los Servicios Telemáticos Inteligentes para el Ciudadano), financiado por el Ministerio de Industria, Turismo y Comercio a través de la Convocatoria de Ayudas de Plan Avanza 2011.

REFERENCIAS

- [1] European Commission, "i2010 e-Government action plan: Accelerating e-Government in Europe for the benefit of all", 2006
- [2] European Commission, "The European e-Government action plan 2011-2015. Harnessing ICT to promote smart, sustainable and innovative government", 2010
- [3] European Commission, "Eurostat e-government statistics: Online interactions of European businesses and citizens with public administrations", 2009
- [4] Kortensniemi, Y., Särelä, M., "Survey of certificate usage in distributed access control", *Computers & Security*, 2014
- [5] Nyström, M.: "PKCS #15 - a cryptographic token information format standard", *USENIX Workshop on Smartcard Technology, WOST'99*, pp. 5-5, 1999
- [6] Instituto Nacional de Estadística, "Encuesta sobre equipamiento y uso de tecnologías de la información y comunicación en los hogares", 2013
- [7] Heichlinger, A., Gallego, P., "A new e-id card and online authentication in Spain", *Identity in the Information Society*, DOI 10.1007/s12394-010-0041-3, 2010
- [8] Ruiz-Agundez, I., Bringas, P.G., "Service authentication via electronic identification cards: Voip service authentication through the DNIe", *SRII Global Conference*, pp.602-607. IEEE Computer Society, 2012
- [9] Olsen, D., Wygant, S., Brown, B., "Electronic survey administration: Assessment in the twenty-first century", *Assessment Update* 16(3), 2004
- [10] Thatcher, J., Waddell, C., "Constructing Accessible Websites", *APress L. P.*, 2003
- [11] United Nations Public Administration Network (UNPAN), "United global nations global e-government survey", 2012
- [12] European Commission, "Eurostat statistic theme 10: Good governance", 2011
- [13] Jefatura del Estado, "Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los servicios públicos", *BOE* núm. 150, 2007
- [14] Ibáñez, J., "Más allá de la sociología: el grupo de discusión: teoría y crítica. Sociología y política", *Siglo Veintiuno de España*, 1979
- [15] Jefatura del Estado, "Ley 39/2006, de 14 de diciembre, de promoción de la autonomía personal y atención a las personas en situación de dependencia", *BOE* núm. 299, 2006
- [16] Net Applications, "Desktop top operating system share trend", <http://marketshare.hitslink.com/operating-system-market-share.aspx> 2015
- [17] W3Schools, "Browser statistics and trends", http://www.w3schools.com/browsers/browsers_stats.asp, 2015
- [18] Marchesini, J., Smith, S., Zhao, M., "Keyjacking: the surprising insecurity of client-side SSL", *Computers & Security* 24 (2), 2005
- [19] Microsoft, "Creating an autorun enabled application", [http://msdn.microsoft.com/en-us/library/cc144206\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/cc144206(VS.85).aspx), 2012
- [20] Microsoft, "Using hardware autoplay", [http://msdn.microsoft.com/en-us/library/cc144210\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/cc144210(VS.85).aspx), 2012
- [21] Microsoft, "Microsoft security intelligence report", Volume 11, 2011
- [22] OPSWAT Market Share Reports, "Antivirus and Compromised Device Report: January 2015", 2015

AC TV Manager: Nuevo sistema de gestión para IPTV basado en software libre

Alain Chas, Jose M. Jimenez, Andrés López-Herreros, Alejandro Canovas, Jaime Lloret
Instituto de Investigación para la Gestión Integrada de Zonas Costeras,
Universitat Politècnica de València
46730 Grao de Gandia (Valencia), Spain.
chaster8@gmail.com, jojiher@dcom.upv.es, andreslohe@doctor.upv.es, alcasol@epsgr.upv.es,
jlloret@dcom.upv.es

Resumen- La evolución de la televisión tradicional al formato IPTV se está produciendo a una gran velocidad, debido a los profundos cambios aparecidos tanto a nivel de sistemas y redes de comunicaciones como de dispositivos. Esta evolución se está haciendo patente en muchos países, y conlleva incluso a la aprobación de regulaciones de neutralidad en la red, como lo ha hecho FCC (Federal Communications Commission). Por todos estos motivos, los nuevos proveedores/suministradores de servicios de IPTV necesitan disponer de herramientas que permitan la gestión de sus contenidos. En este trabajo presentamos una nueva herramienta AC TV Manager, que hemos diseñado, desarrollado e implementado, y que sirve para la gestión de los canales y las emisiones vía streaming que proporciona un servicio de IPTV. Se ha desarrollado en lenguaje Java y sirve como interfaz entre un usuario administrador, VLC y FFmpeg, pues nuestro objetivo principal es que nuestra herramienta no sea propietaria.

Palabras Clave - IPTV, gestión, canales, VLC, FFmpeg

I. INTRODUCCIÓN

El auge de la Televisión vía IP en los últimos tiempos viene acompañado por las grandes opciones que ofrece la tecnología en materia de difusión y emisión de material multimedia. La transmisión de vídeo a través del protocolo TCP/IP comporta la codificación de su información en diferentes capas de este modelo [1]. Ahora bien, dependiendo de la tecnología utilizada para la difusión, existen unos mecanismos para seleccionar diferentes canales dinámicamente, así como para administrar el ancho de banda disponible en el servicio IPTV; estas son algunas de las principales características de la oferta multimedia a través de multicast y que, dependiendo de la tecnología de transmisión usada (red cableada, Wifi, 3G, 4G, WiMAX, etc.) implicará una mayor o menor calidad en la QoE (Quality of Experience) del usuario final [2]. Cada una de estas opciones multicast responde a un estándar y necesita la modulación adaptativa y la codificación para sus emisiones [3].

Uno de los problemas a los que hace frente la oferta multicast de televisión es el llamado “cuello de botella”, que puede originar una petición masiva del mismo; esto se puede evitar multiplexando multicast [4]. Asimismo, la codificación del vídeo es fundamental ya que garantizará una mayor o menor calidad subjetiva final, menor tiempo de transcoding, compatibilidad con navegadores y compatibilidad con framerates en algunos países[5, 6]. La compresión del flujo también es un punto importante en la emisión broadcast, ya que puede solventar en parte problemas de ancho de banda. La combinación de CBR (Constant Bitrate) y VBR (Variable Bitrate) ayudarán a confeccionar un mapa de tramas más

eficaz, así como el número ideal de Intraframes dentro de la secuencia GOP (Group of Pictures). Así es como funcionan actualmente diferentes sistemas de IPTV Multicast a través de proveedores ISP [7].

Como decíamos al principio, el aumento de las tecnologías y las prestaciones en ancho de banda, codificación y protocolos, hacen necesaria la introducción de herramientas, para mejorar la experiencia y las posibilidades del usuario. El resultado es una constante demanda del uso de IPTV así como una transición de la emisión multimedia unicast a las grandes posibilidades del servicio multicast. Esta tecnología es fundamental para el futuro ya que progresivamente va solucionando sus limitaciones de latencia, jitter o multiplexado, y garantiza una QoE [8] mucho mayor que otros servicios unicast o broadcast.

El objetivo del presente trabajo ha sido diseñar, desarrollar e implementar una aplicación que denominamos AC TV Manager. La herramienta que se ha desarrollado, a diferencia de otras propietarias, resuelve problemas de escalabilidad, compatibilidad y velocidad. Permite la emisión en diferentes entornos y la gestión del tráfico de red para emisiones unicast, broadcast y multicast.

La gran apuesta y novedad es su fácil implementación debido a la utilización de programación en Java. AC TV Manager es un programa desarrollado en Java que actúa como interfaz entre el usuario y VLC (Video Lan Client) con el objetivo de facilitar la configuración de la emisión de medios en streaming a través de datagramas IP. El programa funciona como gestor y servidor de streaming de vídeo y audio en red. Nuestro principal objetivo ha sido desarrollar un sistema de gestión de contenidos (CMS), que pueda ser usado en cualquier entorno, pues su funcionamiento se basa en el uso de software libre.

El resto del trabajo se organiza de la siguiente manera: en la Sección II se presentan los trabajos relacionados. En la Sección III se describe la estructura general del sistema de gestión así como su modo de funcionamiento. En la Sección IV se muestran las pruebas realizadas para forzar el tráfico de datos en la aplicación utilizando diferentes protocolos de red. La Sección V presenta las conclusiones y trabajos futuros.

II. TRABAJOS RELACIONADOS

En el mercado hay múltiples sistemas propietarios que son utilizados entre los distribuidores de IPTV.

Concretamente Cisco Systems [9], ha creado una arquitectura para reducir el tiempo de envío y transportar

video de manera eficiente. Cisco IPTV multicast está siendo utilizado para videoconferencias (e-learning) así como video bajo demanda (VoD). El cuello de botella está solucionado con la multiplexación de hasta 50 usuarios simultáneos, momento a partir del cual se necesita una mayor codificación de las tramas de audio y vídeo. La infraestructura de Cisco para este tipo de servicio requiere de servidores de streaming, backbones para multicast a 100, 300, 900 Kbps, una red ACNS (Cisco Application and Content NetworkingSystem) así como un servicio al cliente a través de LAN, Wireless, 3G o WiMax y de VPN para acceso remoto de los trabajadores.

La empresa Exterity propone AvediaServer [10], que es un sistema centralizado de gestión de IPTV. Plantea tres opciones de plataformas y software, que pueden servir en función de su escalado dentro de cualquier ámbito empresarial. Transporta el contenido de cualquier fuente, a través de una red, a cualquier dispositivo sin comprometer la disponibilidad o el rendimiento del sistema IPTV.

AntikTechnology propone Antik IPTV Management Server [11], una solución Software que permite configurarlo todo, desde la gestión de canales hasta la configuración de los STB (Set-Top Boxes) de los clientes. Incluso permite conexiones a aplicaciones de facturación de clientes.

Por otro lado MWAREIPTV presenta IPTV Middleware System [12], que presenta una arquitectura modular para el control total del servicio IPTV. Permite configurar y gestionar varios servicios IPTV desde una consola de administración basada en web central.

Snap.TV ofrece IPTV Content Management System [13], su sistema de gestión de contenidos de IPTV. Permite crear y distribuir infochannels combinando el material local, basado en web, y en directo o diferido, de IPTV.

ENCODEDMEDIA presenta IPTV Manager [14], que proporciona un análisis IPTV remoto, software de monitorización y administración desde Windows. Ofrece una vista de los dispositivos Encoded Media en la red. Permite obtener información en tiempo real del estado de los canales, streams y ancho de banda utilizado, además de información de hardware que incluye el uso de CPU y su temperatura.

Todas las aplicaciones/entornos presentados previamente tienen la característica de ser propietarios de una empresa. Nuestra aplicación se diferencia de todas ellas en que se puede instalar en cualquier equipo con S.O. Windows y actúa sobre software libre (VLC y FFmpeg).

III. ESTRUCTURA GENERAL DEL SISTEMA DE GESTIÓN

AC TV Manager es un programa desarrollado en Java que funciona como interfaz entre el usuario y VLC (Video LanClient) con el objetivo de facilitar la configuración de la emisión de medios en streaming a través de datagramas IP. Este programa permite configurar los horarios y los destinos a los que los medios deben emitirse, de modo que resulte sencillo gestionar un canal de IPTV.

El programa da soporte a los siguientes protocolos: UDP, RTP, RTSP y HTTP, permitiendo además la creación de hasta N canales funcionando simultáneamente. También incluye la opción de una transcodificación en vivo independiente para cada canal, por lo que un conversor, a priori, no resulta necesario. Por otra parte, el software también se ayuda de la funcionalidad del proyecto FFmpeg para labores de importación de medios (obtención de información e imágenes).

En esta sección presentamos la estructura general del sistema, las comunicaciones entre las diferentes aplicaciones de software empleadas, aspectos técnicos de la aplicación, funcionamiento según el protocolo utilizado, funcionamiento lógico y presentación de la propia aplicación AC TV Manager.

A. Esquema gráfico

La Fig. 1 muestra la estructura interna de la aplicación de forma generalizada. En ella pueden observarse los siguientes elementos:

- Aplicación Java: corre sobre la JVM (Java Virtual Machine) de 32 bits. Es la encargada de generar una GUI con la que el usuario interactúe, así como de procesar la información que se envía y recibe de las diferentes aplicaciones en las que se basa el sistema.
- Consola o terminal: intérprete de comandos del sistema operativo (S.O.).
- FFmpeg: colección de software libre que puede grabar, convertir (transcodificar) y hacer streaming de audio y vídeo. Es utilizado, dentro de la aplicación, para tareas tales como obtener información acerca de los vídeos (duración, formato contenedor, códec, etc.), así como para determinar si un determinado fichero del ordenador es vídeo, audio o ninguno de ellos.
- VLC: VLC media player es un reproductor y framework multimedia libre de código abierto, desarrollado por el proyecto VideoLan. Posee la capacidad de streaming, además de una herramienta para la gestión de ésta (llamada VLM, de "Video Lan Manager"); se puede interactuar con ella a través de una interfaz telnet del propio VLC.

Las medias se clasifican en dos tipos:

- vod: estos son usados para vídeo bajo demanda, y no forman parte dentro del alcance de la aplicación. Son medias que selecciona el cliente en el momento que desee, y cuya reproducción puede controlar.
- broadcast: estos medias se acercan a lo que sería un canal de TV, ya que empiezan a emitirse cuando el administrador lo desee ; el cliente los recibe si los solicita, pero sin tener control sobre ellos. La Fig. 2 muestra la declaración de un elemento media de tipo broadcast.

Los schedule son aquellos elementos de tipo programación, que obligan a VLC a realizar determinada acción cuando se alcanza determinada hora prefijada. La Fig. 3 muestra un ejemplo de un elemento schedule.

B. Comunicaciones entre software

A continuación detallamos los tipos de comunicación que se llevan a cabo dentro del esquema general mostrado en la Fig. 1.

La comunicación entre la aplicación Java y el Shell se consigue a través de los servicios que nos provee la clase "Process" de Java, dicha comunicación es un proceso transparente para el usuario. Es una comunicación bidireccional, ya que desde la interfaz gráfica (GUI) se le dice a la consola del S.O. qué instrucción debe ejecutar, y posteriormente se lee la salida que ésta proporciona.

El envío sirve para instanciar el software VLC, hacer llamadas a FFmpeg, solicitar los puertos TCP activos al S.O. y matar procesos VLC. La recepción sirve para leer la respuesta proveniente de FFmpeg y procesarla.

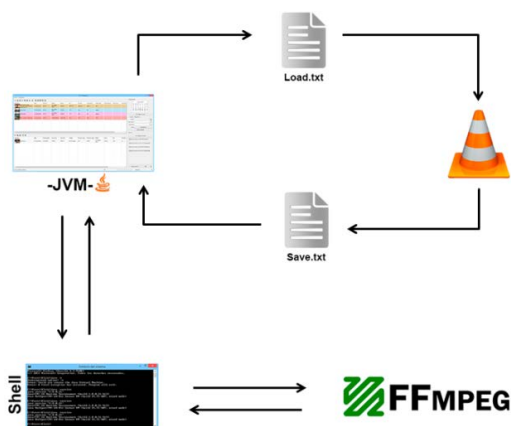


Fig. 1. Estructura general del sistema de gestión

```
new channel1 broadcast enabled
setup channel1 input http://host.mydomain/movie.mpeg
setup channel1 output
#tcp[mux=ts,dst=239.255.1.1,port=5004,sdp=sap://,name="Channel 1*"]
```

Fig. 2. Ejemplo declaración elemento media tipo broadcast en VLC

```
new sched_1schedule enabled
setup sched_1date2013/05/15-12:00:00
setup sched_1 append control channel1 play
```

Fig. 3. Ejemplo declaración elemento media tipo schedule en VLC

```
C:\Users\alain\Desktop\AC TV Manager>ffmpeu -i ./video-1.mp4
Input #0: mov,mp4,m4a,3gp,3g2,mj2, from './video-1.mp4':
Metadata:
  major_brand      : ison
  minor_version   : 512
  compatible_brands: isomiso2avc1mp41
  encoder         : Lavf55.34.101
  Duration: 00:04:49.91, start: 0.000000, bitrate: 2278 kb/s
  Stream #0:0(audio): 0: aac (LC) (mp4a / 0x6134706D), 44100 Hz, 1920x1080
  [SBR] (1: 00R 16:9), 214 kb/s, 23.98 fps, 23.98 tbr, 90k tbn, 47.95 tbc (default)
Metadata:
  handler_name    : VideoHandler
  Stream #0:1(audio): 1: aac (LC) (mp4a / 0x6134706D), 44100 Hz, stereo, flt
  p. 125 kb/s (default)
Metadata:
  handler_name    : SoundHandler
```

Fig. 4. Respuesta de Ffmpeg al Shell

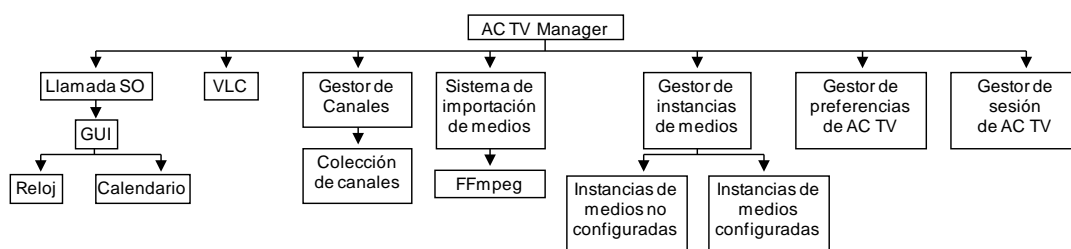


Fig. 5. Módulos del AC TV Manager

La comunicación entre Shell y FFmpeg es un proceso automático y permite la solicitud y recepción de información acerca de los medios importados al sistema. La Fig. 4 muestra un ejemplo de la respuesta dada por el software FFmpeg tras solicitar información acerca de cierto vídeo, a través de la línea de comandos. La respuesta aporta datos tales como duración, tipo de códec, etc.

La comunicación entre Java y VLC tiene como prerequisites el que el VLC debe haber sido instanciado con interfaz telnet activa y se debe haber establecido una conexión telnet. A partir de tal conexión se puede dar órdenes a la herramienta VLM del VLC. Básicamente, los comandos que se utilizan y que interpretará la herramienta VLM son "load" para cargar la configuración del fichero load.txt y "save" para obtener una copia (save.txt) de la configuración guardada y corroborar que VLC está cumpliendo con su cometido.

C. Aspectos técnicos de la aplicación

El programa se compone de varias partes, que pueden verse en forma de módulos, tal como muestra la Fig. 5.

Dentro de "llamada al S.O." tenemos el modulo "GUI", que es la interfaz gráfica de la aplicación, mediante la cual interactuará el usuario. Está compuesta por widgets de la librería gráfica SWT (una librería open-source para Java creada por el proyecto Eclipse), los cuales son llamadas directas a widgets gráficos intrínsecos del S.O. Entre los elementos gráficos destacamos "Reloj" y "Calendario". El "Reloj", muestra la hora del sistema en el que se esté ejecutando la aplicación. También es una herramienta base para la gestión horaria de las emisiones, ya que en ella se basará VLC para emitir, y ayuda a detectar qué instancias de medios ya han sido emitidas y permite, por tanto, la limpieza automática. El "Calendario" es otra pieza fundamental, permitiendo al usuario la selección del día en que se emitirá el medio a configurar. Además, permite una rápida navegación para la visualización de la programación creada.

AC TV Manager delega las funciones de streaming al software VLC. El módulo "VLC" es el encargado de las funciones de streaming.. Se comunica con él vía Telnet para cargarle la nueva configuración horaria y de emisión cada vez que el administrador realiza un cambio en el programa. Desde el panel de preferencias de la aplicación puede asignarse la ruta del ejecutable "vlc.exe" deseada, por lo que el sistema podría utilizar nuevas versiones del software VLC a medida que estas vayan apareciendo.

El sistema permite anunciar el contenido emitido a través de anuncios SAP (Session Announcement Protocol) que VLC envía automáticamente a la dirección multicast 224.2.127.254. Estos anuncios se producen sólo con el uso de los protocolos UDP y RTP. En el caso concreto de RTP, además, se hace a través de un formato descriptivo acorde a SDP (Session Description Protocol). Esta información SDP también puede ser guardada como fichero para un modelo de difusión diferente.

El módulo "Gestor de canales" es el conjunto de herramientas que nos permite añadir vías de salida de stream. El programa permite la creación de tantos como desee el administrador. El canal se compone de:

- Número identificador.
- Nombre.
- Protocolo (UDP, HTTP, RTP o RTSP).
- IP destino (sólo afectará sobre los protocolos UDP y RTP; HTTP y RTSP actúan como servidor de stream bajo demanda, por lo que la IP será la del propio ordenador donde esté corriendo la aplicación).
- El modelo de encapsulación (o muxer) del stream. Los modos disponibles son "ts" (transportstream), "ogg", "webm", "asf" y "mkv". El único muxer válido para todos los protocolos es "ts" (transport stream), mientras que los demás se limitan a funcionar con HTTP.

La extensión es un campo optativo, es el conjunto de caracteres que acompañarán a la dirección por defecto del

stream. Por ejemplo, un canal UDP con IP destino 239.0.0.1 y puerto 7050; y con la extensión "stream-pruebas", su dirección de acceso quedaría de la siguiente forma: `udp://239.0.0.1:7050/stream-pruebas`. Cada canal posee, además, la posibilidad de transcodificación en vivo, permitiendo diferentes parámetros de configuración.

El programa permite la superposición de un logotipo para cada canal de emisión, con la restricción de que debe activarse la transcodificación en vivo. Éste logo se posiciona en la esquina inferior derecha de la pantalla.

El módulo "Sistema de importación de medios" es el conjunto de herramientas que el programa ofrece para la importación de los medios. El botón destinado para ello abre un explorador de Windows para la selección de los archivos a importar. También se puede hacer mediante "Drag&Drop" desde cualquier ventana de exploración del S.O.

El sistema ofrece la posibilidad de emitir contenido de fuentes externas. Para ello debe crearse una instancia especial y escribir un modelo de fuente que VLC pueda reconocer. Ejemplos de fuentes válidas son `udp://@239.255.1.1:6050` o `http://dominio.com/video.avi`, de modo que resulta posible usar contenido proporcionado por terceros. Cada vez que se importan archivos, estos pasan por un "filtrado" para determinar si son medios de audio o vídeo válidos y obtener, de paso, sus propiedades relevantes (duración, bitrate, resolución, fps, formato, códec de audio y códec de vídeo). El filtrado se crea a través de la información devuelta por una llamada al software `ffprobe` (parte de `FFmpeg`). Finalmente, el módulo "FFmpeg" se encarga de obtener un "thumbnail" representativo del vídeo, antes de crearse la entrada de tabla.

El módulo "Gestor de instancias de medios" gestiona las referencias a los medios almacenados en el equipo, discos duros, red, etc. Las instancias vienen representadas por una entrada de tabla, y son los elementos que deben ser configurados (mediante asignación de fecha y hora) para una correcta emisión del medio por alguno (o varios) canales. El gestor de instancias se encarga de mover éstas de la tabla de no-configurados a la tabla de configurados si han sido preparadas para su emisión. Las instancias están configuradas cuando tienen una hora de inicio asignada, cuando tienen una hora de final asignada o cuando están asignadas a uno (o varios) canales.

El módulo "Gestor de preferencias" se encarga de aplicar cambios globales en ciertos comportamientos de la aplicación. Las preferencias no se almacenan en una sesión (es decir, no son almacenadas en el fichero creado al guardar una sesión), sino que son intrínsecas a la aplicación en sí, y se mantienen al cerrar y volver a abrir el programa. Son ejemplo de ello la ruta alternativa a VLC y número asignado al TTL (Time To Live) de los paquetes de los streams.

El módulo "Gestor de sesión" se encarga de transcribir toda la información de la sesión a un fichero. Este fichero se puede guardar manualmente, aunque el programa provee un sistema de auto-guardado para evitar pérdida de información ante un cierre inesperado. Se llama sesión al estado en que se encuentra el programa en determinado momento. Esta incluye las instancias de medios no configuradas, las instancias de medios configuradas y los canales y parámetros de transcoding de cada uno.

La ruta de autoguardado es aquella en la que se halle la aplicación, dentro de la carpeta "autoSave". El gestor de sesión también es el encargado de volver a crear la sesión ante la apertura de un fichero previamente guardado.

D. Funcionamiento dependiendo del protocolo utilizado

Al emplear AC TV Manager como servidor de streaming, cuando utiliza los protocolos HTTP y RSTP, el cliente inicia la comunicación enviando una señal de "Solicitud de stream" y el "AC TV Manager" le responde enviando el stream demandado en modo unicast. En el caso de utilizar los protocolos UDP o RTP, "AC TV Manager" envía streams hacia los clientes (puede ser tanto unicast como multicast).

E. Funcionamiento lógico

El algoritmo de funcionamiento de AC TV se muestra en la Fig. 8. Tras arrancar AC TV Manager, se comienza con "Iniciar la emisión". Seguidamente pulsamos sobre "Importar medios". Para esto deberemos pasar a "Seleccionar un medio". Una vez seleccionado lanzaremos una "Llamada a FFmpeg" para conocer si el fichero que vamos a importar es adecuado o no. En el caso de que el fichero no fuera válido, retrocederemos al paso "Importar medios", en el caso contrario se crea una "Instancia de medio". Si queremos enviar más medios en el stream, volveremos de nuevo a "Importar medios" y repetiremos la secuencia descrita previamente. Si hemos importado todos los medios, pasaremos a verificar que tenemos creado el canal por el que vamos a emitir el stream, en el caso de que ya estuviera creado continuaremos con "Añadir instancia de medio a canal". En caso de no tener creado el canal accederemos a "Crear canal" y posteriormente a "Configurar canal", desde aquí ya vamos a "Añadir instancia de medio a canal". El siguiente paso es "Hacer corrección horaria de la planificación de las instancias de medio añadidas al canal", cuando ya se ha realizado seguiremos con "Actualizar la planificación horaria en VLC". Llegado a este punto, si tuviéramos que "Añadir más instancias de medios al canal" volveríamos al paso de "Añadir instancia de medio a canal" y repetiríamos la secuencia. Si ya las hemos añadido todas las instancias de medios, llegaremos al punto en el que decidiremos si queremos añadir más canales. Si queremos añadir más canales volveremos al paso "Crear canal", en el caso contrario decidiríamos si queremos "Importar nuevos medios". Si quisiéramos importar nuevos medios volveríamos al paso "Importar medios" y seguiríamos todos los pasos de nuevo, si ya los hubiéramos importado todos finalizaríamos.

F. Aplicación AC TV Manager

El entorno de la aplicación está compuesta por: la interfaz inicial, importación de medios, información de los medios, configuración de las instancias de medios, configuración de los canales, inicio de emisión, monitorización, Drag&Drop, intersección de medios, gestión de la sesión y limpieza automática de instancias.

La "Interfaz inicial" es escalable, aunque se recomienda utilizarla en un monitor con resolución mayor a 960x640 píxeles. En la interfaz nos encontramos con las siguientes zonas que se identifican en la Fig. 9:

1. Menú superior (Zona 1). Ofrece opciones generales tales como guardado y apertura de la sesión, o la gestión de preferencias de la aplicación.
2. Barra de Herramientas 1 (Zona 2). Barra de herramientas relativa a la tabla 1, con utilidades que afectan a las instancias de medios que se encuentran en ésta.
3. Tabla 1 o tabla de instancias no configuradas (Zona 3). Es la tabla que almacena las instancias de medios que aún no han sido configuradas para su emisión.

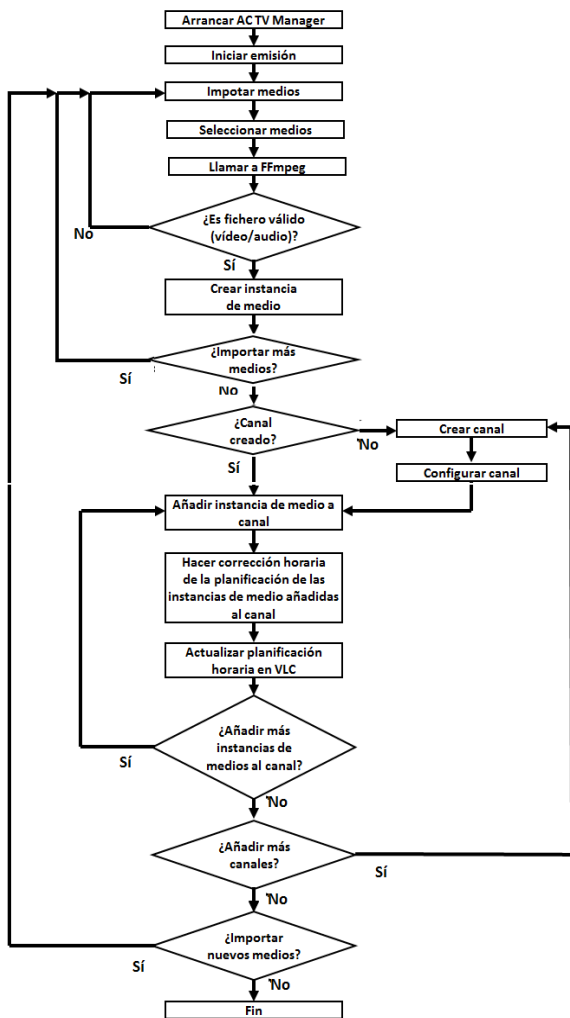


Fig. 8. Algoritmo de funcionamiento AC TV Manager

4. Panel de configuración de instancias (Zona 4). Panel con opciones de configuración de la instancia de medio seleccionada (fecha de inicio, hora, etc.).
5. Barra de Herramientas 2 (Zona 5). Barra de herramientas relativa a la tabla 2, con utilidades que afectan a las instancias de medios que se encuentran en ésta.
6. Tabla 2 o tabla de instancias configuradas (Zona 6). Es la tabla que almacena las instancias de medios que ya han sido configuradas para su emisión.
7. Panel de emisión (Zona 7). Es el panel que permite la adición o borrado de canales, así como su configuración. Además, posee los botones de inicio y parada generales.
8. Reloj, etiquetas y barra de progreso (Zona 8). Sirven para monitorización de procesos del programa (tales como la inicialización de la emisión o la importación de archivos). Primero se debe hacer la "Importación de medios".

Existen dos formas de hacerlo: mediante el Botón Importación, en la barra de herramientas 1 o mediante Drag&Drop desde cualquier ventana de exploración de Windows a la tabla 1. El programa filtrará cuáles son válidos y los incorporará a la sesión, pues es capaz de importar todo aquello que las librerías del proyecto FFmpeg sean capaces de interpretar como vídeo y/o audio. Tras la importación, se crean instancias de medios, una por cada vídeo y/o audio importado. Estas instancias son referencias al fichero de vídeo o audio original, y se presentan en la tabla 1, tal como muestra la Fig. 10. La "Información de los medios" esta presentada en las tablas, así como sus propiedades de emisión. Los parámetros que se muestran repartidos en columnas son: Nombre (acompañado de un thumbnail extraído del medio), Ruta, Duración del medio, Bitrate, Escala, fps, Formato, Códec audio, Códec vídeo, Fecha emisión, Hora inicio, Hora final, Canales. Se puede saber la información extendida (dada por el software FFprobe del proyecto FFmpeg) de un determinado medio haciendo click sobre su instancia en la tabla y pulsar el botón Información.

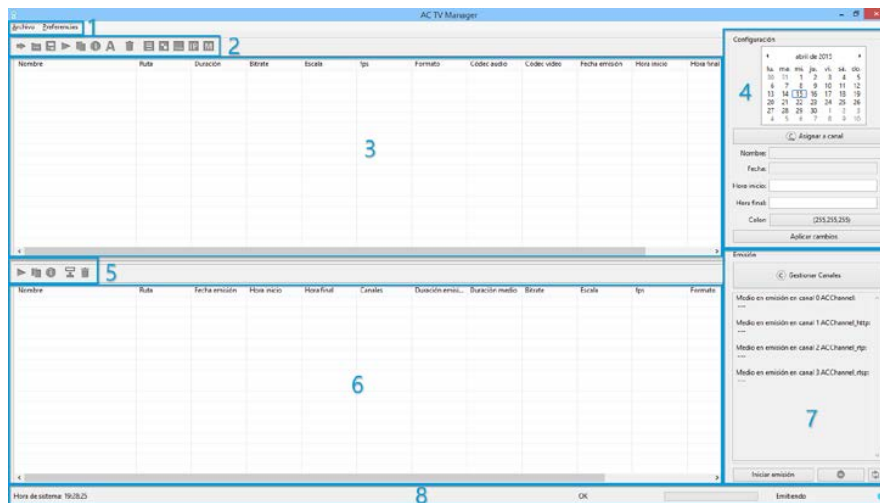


Fig. 9. Interfaz inicial

Nombre	Ruta	Duración	Bitrate	Escala	fps	Formato	Códec audio	Códec vídeo	Fecha emisión	Hora inicio	Hora final
bbb_sunflower_1080p_30f...	C:\Users\Alain...	0:10:34	3,481424 Mbit/s	1920x1080	30/1	mov	mp3	h264			
bueno.mp4	C:\Users\Alain...	0:04:52	369.372000 Kbit/s	320x240	15/1	mov	aac	mpeg4			
Black Label Society Concrete Jungle [OFFICIAL VIDEO].webm	C:\Users\Alain...	0:03:38	753.442000 Kbit/s	448x336	2997/125	matroska	vorbis	vp8			
Big_Fish.avi	C:\Users\Alain...	2:00:01	817.627000 Kbit/s	520x272	25/1	avi	mp3	mpeg4			

Fig. 10. Importación de medios

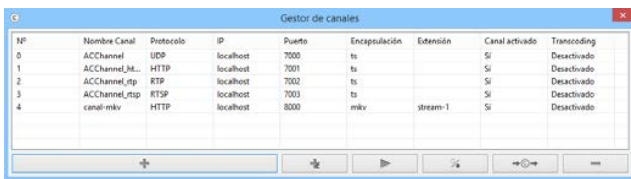


Fig. 11. Configuración de canales

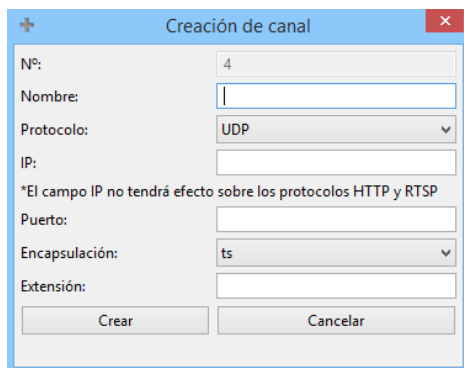


Fig. 12. Creación de nuevo canal

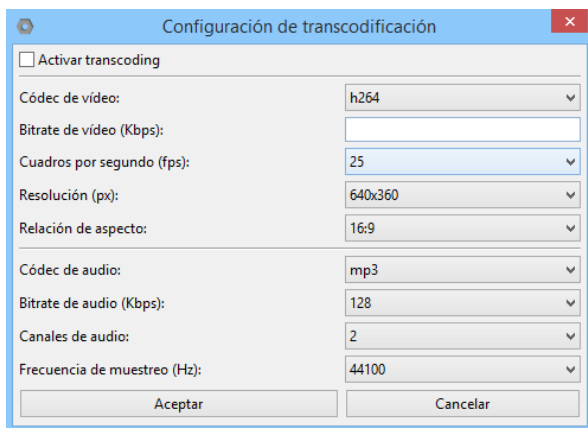


Fig. 13. Configuración de la transcodificación



Fig. 14. Visor de listado

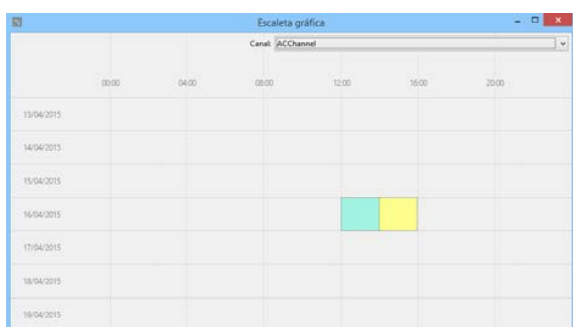


Fig. 15. Visor de escala gráfica

Para la "Configuración de las instancias de medios" se debe pulsar en una entrada de la tabla 1. Se observará que el campo de Nombre del panel de configuración se actualiza automáticamente (ver área 4 en Fig. 10). Seguidamente se puede configurar el horario y canal de emisión.

La creación y "Configuración de canales" de emisión es el segundo punto clave, junto con la configuración de instancias de medios previamente vista. Para acceder al panel de gestión de canales se presiona sobre el botón "Gestionar Canales" del panel Emisión, que abrirá la ventana mostrada en la Fig. 11. Esta ventana muestra una tabla con el listado de canales creados y una botonera inferior con las siguientes funcionalidades: Adición de canal, Copia de canal, Previsualización de canal, Activación/Desactivación, Transcoding y Eliminar canal. La Fig. 12 muestra la ventana de creación de un nuevo canal. En el caso de ser necesaria la transcodificación del stream se podrá realizar usando la "Configuración de la Transcodificación", cuya ventana de puede observar en la Fig. 13. Se debe activar "Inicio de la sesión", aunque ya estén configuradas las instancias de medios y asignados los canales. Para iniciar la emisión, se debe pulsar el botón "Iniciar emisión" del panel Emisión. Si todo ha funcionado correctamente, las etiquetas de la parte inferior de la aplicación habrán adoptado el aspecto que muestra la Fig. 14. Al iniciar la emisión, el programa llama a una instancia VLC que funciona en segundo plano, sin interfaz gráfica a esta instancia, en VLC se carga la configuración de emisión de medios realizada hasta el momento, así como, posteriores cambios y configuraciones que sean realizadas (todo de forma automática y transparente para el usuario). VLC actuará como servidor, realizando streaming y transcodificación. La detención de la emisión provoca que finalice la instancia VLC funcionando en segundo plano. Para ello, se debe presionar el botón "Detención de la emisión" del panel Emisión.

Para realizar la "Monitorización", la aplicación ofrece varios modos de ver la información configurada sobre la instancia VLC que actúa como servidor. Los tres modos principales son: Visor de listado, que además permite cambiar fácilmente el orden de emisión de los medios (Fig. 14), Visor de escala gráfica (Fig. 15) y Visor VLM, que muestra el código cargado en VLC (Fig. 16).

Dado que puede resultar tedioso configurar manualmente el horario de inicio y parada de cada instancia de medio, así como la asignación individual de canales por los que se emitirá, el programa ofrece funciones de Drag&Drop para un flujo de trabajo óptimo. Para poder configurar un medio a continuación de otro, y con los mismos canales destino, según muestra la Fig. 17 se puede arrastrar cualquier instancia de medio de la tabla 1 y depositarla sobre cualquier instancia de medio de la tabla de instancias configuradas (flecha 1) o del visor de listado (flecha 2).

Mediante la "Intersección de medios" se evitan conflictos de emisión ya que no se permite que dos medios compartan canal de salida al mismo tiempo; es por ello que el programa no permite realizar ninguna configuración horaria o de canal que provoque tal situación. Si el usuario intenta realizar una configuración de ese tipo, el programa no aplicará el último cambio y avisará con un mensaje como el de la Fig. 18.

Denominamos sesión al estado en que se encuentra el programa en un determinado momento, la "Gestión de la sesión" permite guardar o abrir una sesión previamente guardada, en cuyo caso se cerraría la sesión actual.

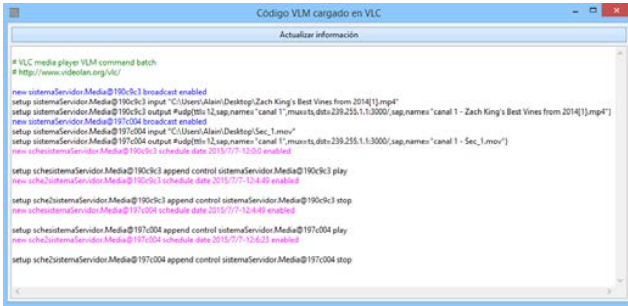


Fig. 16. Visor VLM

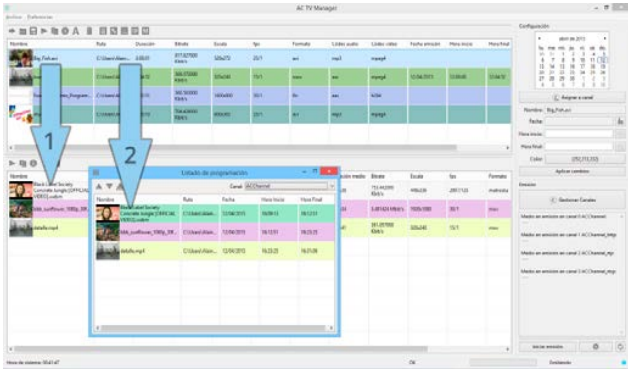


Fig. 17. Drag&Drop

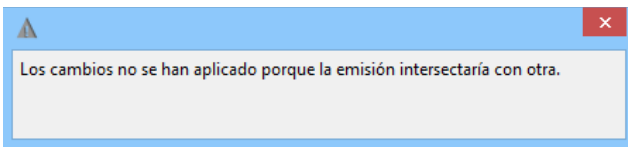


Fig. 18. Mensaje de error en la Intersección de medios

La aplicación realiza de forma automática una "Limpieza de instancias" de las instancias de medios ya emitidos. De esta forma se evitan conflictos de emisión y se mantienen las tablas de la aplicación más limpias y ordenadas.

IV. TEST DE RENDIMIENTO

Para realizar el test de rendimiento, hemos utilizado la topología mostrada en la Fig. 19, compuesta principalmente por el equipo donde se instaló AC TV Manager, los equipos clientes donde se recibía el stream, y todos los dispositivos de red y cableado necesarios para la interconexión. Como se puede observar, los clientes y el sistema gestor se encuentran ubicados en redes diferentes. La transmisión a los clientes se realiza de forma inalámbrica dentro de la red en la que se encuentran ubicados, mientras que el servidor se conecta con cable. En cada uno de los clientes se instaló un cliente VLC para la recepción del vídeo y la aplicación Wireshark para poder obtener información de la transmisión recibida.

En la primera medición utilizamos un vídeo que tenía una tasa de bits de 1355 kbps (1163 vídeo + 192 audio) y una resolución de 1280x720 y 25 fps (frames por segundo). Se efectuó la transmisión con cada uno de los protocolos seleccionados por la herramienta (HTTP, RTP y UDP). Los resultados se muestran en la Fig. 20. Éstos fueron obtenidos utilizando la herramienta Wireshark. Se puede observar que la transmisión mediante RTP es la que utiliza mayor ancho de banda (exceptuando entre los segundos 15 y 16), en los que el ancho de banda consumido es inferior al de UDP. Mientras que cuando utilizamos HTTP o UDP, aunque la transmisión inicial de UDP utiliza un mayor ancho de banda, rápidamente pasan a tener anchos de banda similares.

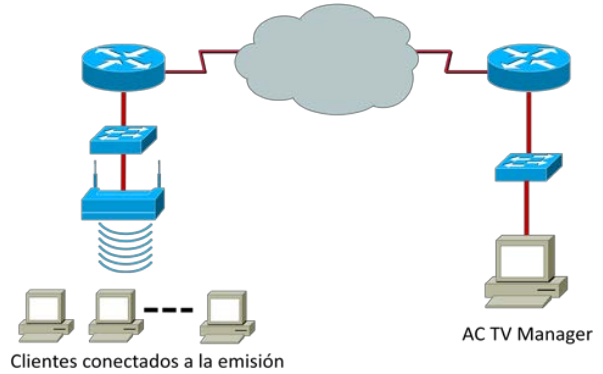


Fig. 19. Topología base de transmisión

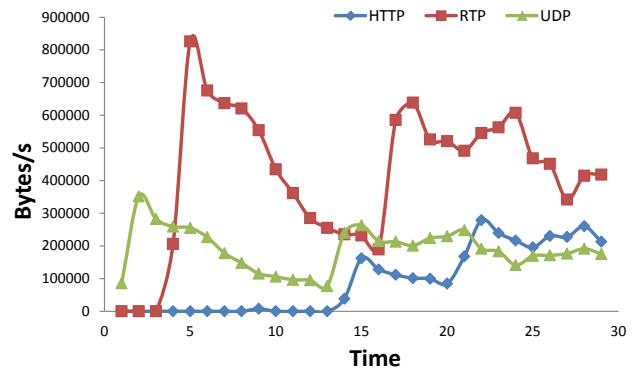


Fig. 20. Comparativa de captura de tráfico en los tres protocolos a 1355 Kb/s

El pico más alto lo alcanza RTP con más de 800000 Bytes por segundo. La media de bytes por segundo es mayor en RTP, siendo UDP el menor.

En la segunda medición utilizamos un vídeo que tenía una tasa de bits de 2499 kbps (2307 vídeo + 192 audio), una resolución de 1280x720 y 25 fps. Se efectuó la transmisión sobre cada uno de los protocolos seleccionados por la herramienta (HTTP, RTP y UDP). Los resultados obtenidos se muestran en la Fig. 21. Dichos resultados fueron capturados utilizando la herramienta Wireshark. Se puede observar que la transmisión mediante RTP es casi idéntica a la transmisión UDP, mientras que la transmisión HTTP utiliza un ancho de banda notablemente inferior a las anteriores. Observamos que el pico máximo de Bytes transmitidos en un instante por HTTP es un valor aproximado a los 572000, sin embargo los valores de RTP y UDP superan los 783000 y 725000 respectivamente. El valor promedio es superior a 322000 Bytes/s.

Para realizar la tercera medición, generamos un vídeo con una tasa de bits de 3078 kbps (2886 vídeo + 192 audio), una resolución de 1280x720 y 25 fps. Como se puede observar en la Fig. 22, los bytes recibidos por segundo son similares al emplear cualquiera de los tres protocolos HTTP, RTP o UDP. Solo se observa un aparente retardo en el tiempo inicial sobre el protocolo UDP. Este retardo inicial podría hacernos llegar a la conclusión de que el cliente final estaba teniendo una mala QoE del vídeo recibido, pero no era el caso, pues gracias a los buffers de VLC el cliente final observaba correctamente el vídeo transmitido. El pico máximo ha sido medido en UDP, pero no había gran diferencia con los otros. La media de los 3 protocolos ha variado entre 425.000 aproximadamente y 343000, siendo el valor más alto el de HTTP y el más bajo el de UDP.

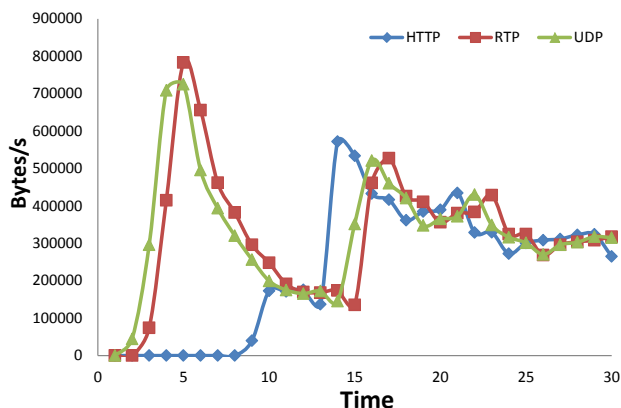


Fig. 21. Comparativa de captura de tráfico en los tres protocolos a 2499 Kb/s

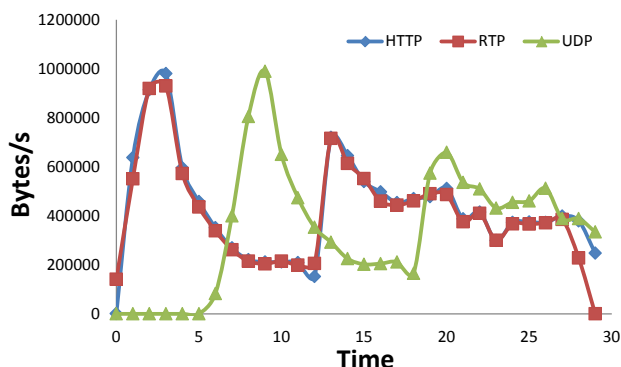


Fig. 22. Comparativa de captura de tráfico en los tres protocolos a 3078 Kb/s

V. CONCLUSIONES

En este artículo hemos presentado la aplicación AC TV Manager desarrollada en Java. La aplicación sirve para gestionar la emisión de diferentes medios multimedia mediante streaming. La aplicación se ha diseñado, desarrollado e implementado para que funcione utilizando aplicaciones de software libre muy conocidas y ampliamente utilizadas como son VLC y FFmpeg. Como resultado disponemos de esta aplicación que funciona como sistema de continuidad televisiva de IPTV.

En el estudio de rendimiento hemos puesto a prueba las posibilidades de la aplicación utilizando diferentes protocolos (HTTP, RTP y UDP) para envío de tramas multimedia. El comportamiento es muy diferente según el protocolo utilizado. En HTTP hemos observado que la cabecera MIME realiza un cambio en el servidor, lo que ralentiza el proceso de emisión. Por tanto, el protocolo HTTP no es el más adecuado para la emisión en AC TV Manager. Por otro lado, el protocolo RTP empieza con un pequeño retardo al inicio, aparece el audio e inmediatamente el vídeo, el cual lleva un ligero retardo y aumenta el delay y el jitter provocando errores de pixelado y de movimiento. El protocolo que mejor se comporta en esta aplicación es el UDP donde no aparece ningún error de transmisión ni de imagen, además, la velocidad es mayor a la de HTTP y RTP. Se ha realizado la prueba de emisión para más de 5 dispositivos simultáneamente obteniendo un excelente resultado, ya que la aplicación ha respondido de manera satisfactoria y ha generado un unicast a modo de Video on Demand (VOD) a cada equipo de manera personalizada y sin ningún tipo de fallo, es decir, se ha programado una agenda multimedia para cada uno de los equipos de manera

simultánea obteniendo una fluidez muy alta gracias a la programación de la herramienta en Java.

Una de las grandes ventajas de la herramienta frente a las propietarias existentes es que se trata de una herramienta colaborativa que permite cualquier número de emisiones unicast a diferentes equipos de manera programada y con una fluidez muy alta, el límite es el ancho de banda disponible para el servicio. Otra ventaja respecto a herramientas propietarias es su fácil implementación de código en otras plataformas y otros equipos, lo que permite una gran compatibilidad en una futura apuesta comercial para servicios multimedia unicast, broadcast y multicast.

La previsión para nuestro trabajo futuro es mejorar la herramienta, diseñándola e implementándola para entornos de trabajo colaborativo. Se intentará integrar con las plataformas modernas de transporte de contenidos multimedia IMS para que se puedan publicar los contenidos en la red de un proveedor de servicios. Además, el principal propósito es utilizarla en investigaciones futuras en las que se necesite un sistema de gestión de contenidos, pues de esta forma disponemos de la herramienta gratuita, no propietaria, y que puede ser modificada en función de nuevas necesidades.

VI. REFERENCIAS

- [1] D. Banodkar, K. K. Ramakrishnan, S. Kalyanaraman, A. Gerber, O. Spatscheck, "Multicast Instant Channel Change in IPTV Systems", 3rd IEEE International Conference on Communication Systems Software and Middleware and Workshops, 2008. pp. 370-379, January 2008.
- [2] Miguel García, Jaime Lloret, Miguel Edo, Raquel Lacuesta, IPTV Distribution Network Access System Using WiMAX and WLAN Technologies, Proceedings of the 4th edition of the UPGRADE-CN workshop on Use of P2P, GRID and agents for the development of content networks, Munich (Alemania), 11-13 de Junio 2009
- [3] W. Poon, K. Lo, J. Feng, "Performance Study for Streaming Layered Encoded Videos in Broadcast Environment", Information Technology and Applications, 2005. ICITA 2005. Third International Conference on, vol. 2, IEEE, pp. 722-727, July 2005.
- [4] P. C. Ng, S. C. Liew, C. Lin, "Voice over Wireless LAN via IEEE 802.16 wireless MAN and IEEE 802.11 wireless distribution system", Wireless Networks, Communications and Mobile Computing, 2005 International Conference on, vol. 1, IEEE, pp. 504-509, June 2005.
- [5] A. Lopez-Herreros, A. Canovas, J. M. Jimenez, J. Lloret. "A New IP Video Delivery System for Heterogeneous Networks using HTML5". IEEE ICC, London, June 11, 2015.
- [6] Ruth Herrero Doñate, Jaime LloretMauri, Jose Miguel Jiménez Herranz, Estudio de Códecs de Compresión MPEG-4 para su Aplicación a la Videovigilancia, XX Simposium Nacional de la Unión Científica Internacional de Radio (URSI2005), Gandia (Valencia), 14-16 de Septiembre de 2005
- [7] K. Imran, M. Mellia, M. Meo, "Measurements of Multicast Television over IP", Local & Metropolitan Area Networks, 2007. LANMAN 2007. 15th IEEE Workshop on, IEEE, pp. 170-175, June 2007.
- [8] J. M. Jimenez, A. Canovas, A. Lopez-Herreros, J. Lloret, "New algorithm to improve QoE of IPTV service customers", En CSSMA-06: Multimedia traffic and quality modelling, June 11, 2015.
- [9] Deployed IPTV for Global Communications. En: http://www.cisco.com/web/about/ciscoitwork/downloads/ciscoitwork/pdf/cisco_it_case_study_iptv.pdf
- [10] AvidiaServer - Centralized IPTV Management System. En: <http://www.exterity.com/products/avidiasever.html>
- [11] IPTV Management Server. En: <http://www.antiktech.com/en/streamingbripty-headend/management/iptv-management-server>
- [12] IPTV Middleware system. En: <http://iptvmiddleware.com/iptv-middleware-system/>
- [13] IPTV Content Management System. En: http://www.snap.tv/technology/IPTV_cms.html
- [14] IPTV Manager. En: <http://www.encodedmedia.com/iptv-manager/>

Diseño SDN/NFV con retardo balanceado entre orquestador, controladores y elementos de red

Aurelio Javier García, Cristina Cervelló-Pastor, Yury Jiménez
 Departamento de Ingeniería Telemática,
 Universitat Politècnica de Catalunya,
 Esteve Terradas, 7, 08860, Castelldefels, Barcelona.
 aurelioj.garcia@entel.upc.edu, cristina@entel.upc.edu, yury.jimenez@entel.upc.edu.

Resumen—El eje central de esta investigación se concreta en el desarrollo de una metodología específica que consigue seleccionar la ubicación óptima de los nodos de control en las redes SDN y NFV. El objetivo es solventar las problemáticas generadas por los retardos de propagación entre estos nodos y el resto de los elementos de la red. El procedimiento que se expone está formado por una serie de cálculos en torno a las características de los nodos y sus enlaces, así como por los criterios de ordenación de dichos nodos (a partir de los valores de sus características), y por la elección de la ubicación más eficiente de los nodos de control. El criterio de ordenación está relacionado con el objetivo deseado para el retardo. La evaluación de esta metodología nos muestra la relevancia del criterio de ordenación utilizado, en especial cuando el número de estos elementos de red aumenta.

Palabras Clave—Redes, SDN, NFV, Orquestador, Controladores, Retardo

I. INTRODUCCIÓN

SDN (Software-defined Networking) [1] y NFV (Network Functions Virtualization) [2] son técnicas fundamentales en el desarrollo de la Internet del Futuro que van a permitir dar respuesta a la demanda creciente de tráfico, optimizando a la vez la inversión necesaria (CAPEX) y los costes de operación y mantenimiento (OPEX), lo que facilitará nuevas oportunidades de negocio.

En una arquitectura tradicional de red el control se encuentra distribuido entre los dispositivos de red (switches/routers), los cuales se encargan de encaminar los datos entre usuarios. El administrador de red es el que establece la forma en la que se encaminan los datos en cada enrutador. Por el contrario, en la arquitectura SDN, los planos de control y de datos están desacoplados, de forma que la inteligencia de red se halla centralizada y la infraestructura de red subyacente se independiza de las aplicaciones y servicios (Fig. 1). Ello facilita la virtualización de la red y, además, se consigue un mayor nivel de programación, automatización y control de la red, permitiendo una mayor escalabilidad, flexibilidad y adaptación a cambios. Las características de SDN permiten que sea una arquitectura de red adecuada para la naturaleza dinámica y de alto consumo de ancho de banda de las aplicaciones actuales.

La arquitectura de red SDN propuesta en la Fig.1 incluye tres elementos principales: aplicaciones de control de red, controladores de rutas y los elementos de red que reenvían datos. Una aplicación de control de red se encarga de comunicar las necesidades y los comportamientos deseados al controlador de rutas. Este controlador es una entidad lógica de control encargada de traducir las peticiones de la aplicación de control a los elementos de red. A su vez, un elemento de red es un

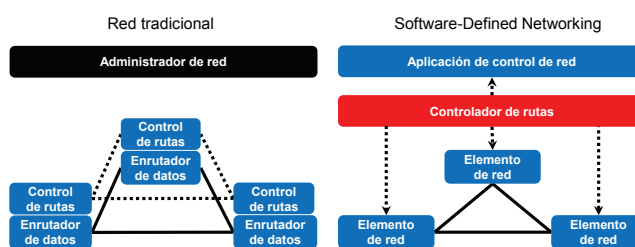


Fig. 1. Red tradicional vs. SDN.

componente lógico que reenvía datos según las tablas de flujos programadas por el controlador que tiene asignado.

Por otra parte, NFV se basa en tecnologías relacionadas con la virtualización. En este contexto, la virtualización de las funciones de red (*Virtualized Network Function* o VNF) tiene como objetivo implementar funciones de red específicas que se ejecutan en una o más máquinas virtuales configuradas sobre servidores, en lugar de funcionar sobre dispositivos de red especializados. Una aplicación de NFV es la construcción de servicios de red mediante la creación bajo demanda de instancias VNF, siendo la capa de orquestación la encargada de la monitorización, reparación y facturación de las VNFs.

SDN es un concepto relacionado con NFV aunque tratan ámbitos diferentes. SDN se centra en la separación de la capa de control de la red, permitiendo así la creación de redes virtuales. NFV se centra en la virtualización de las funciones de la red, para permitir la migración de un esquema de red basado en nodos hardware específicos a otros de propósito general. En la arquitectura NFV se incluye un orquestador que permite desplegar escenarios de red gestionando la infraestructura virtualizada.

El funcionamiento de la arquitectura combinada de ambas tecnologías que se muestra en la Fig. 2 es el siguiente [3]. El orquestador solicita a los controladores crear una red virtual para un cliente. Los controladores proporcionan conectividad a todas las VMs (*Virtual Machines*) de la red virtual bajo la dirección del orquestador. El orquestador tiene visión de las diferentes plataformas de la red y puede monitorizar las VMs en cada servidor y el estado de sus recursos (memoria y CPU). Si una VM alojada en un servidor tiene un mayor número de solicitudes, el orquestador puede asignarle más recursos o crear una imagen suya en una nueva VM.

Los proveedores de infraestructura pueden beneficiarse del alquiler de segmentos de red mediante el uso de redes SDN & NFV entre los proveedores de servicio y los usuarios

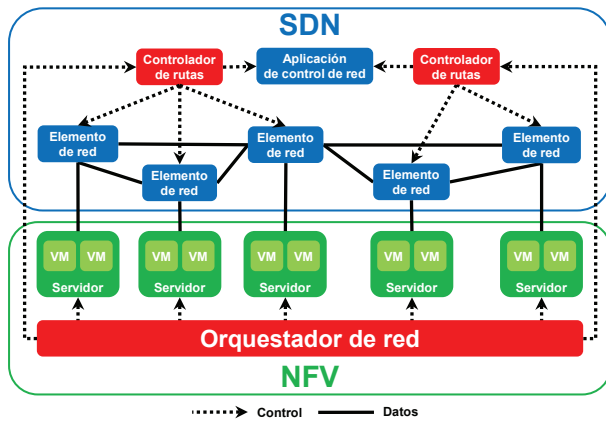


Fig. 2. Arquitectura SDN & NFV.

finales. Estos usuarios, en concordancia con los proveedores de servicio, pueden exigir una serie de requisitos para su servicio, como por ejemplo, un retardo máximo entre determinados nodos de la red asignados a la red del usuario. La variedad de estos tipos de servicios existentes lleva a un desafío a los proveedores de servicio, teniendo que hacer cumplir unos razonables SLAs (Software Level Agreements) para no tener sanciones por el incumplimiento de la QoS (Quality of Service).

La arquitectura combinada presenta dos tipos de nodos: aquellos que envían instrucciones de control, definidos como nodos especiales, y los considerados normales (aquellos nodos que las reciben). Como nodos especiales tenemos el orquestador y los controladores de red y, como nodos normales, el resto. La información de control que envían los nodos especiales tiene que llegar con el menor retardo posible a los nodos normales, pero evitando el incremento del retardo en el reenvío de datos entre nodos normales.

El problema de la localización de controladores en SDN fue propuesto inicialmente por Heller et al. [4], con el objetivo de minimizar la latencia de propagación. En este artículo se utiliza el algoritmo K-center para minimizar la máxima distancia entre los controladores y los nodos de la red. Otras investigaciones consideran objetivos adicionales como la carga entre controladores [5], robustez frente a fallos de nodos y enlaces, y balanceo de carga en el plano de control [6][7]. Sin embargo, muchas de estas métricas compiten entre sí, siendo necesario hallar algoritmos sencillos que den una solución orientativa al problema de la localización de los controladores.

En este trabajo se presenta una metodología sencilla para seleccionar la ubicación de los nodos especiales en una arquitectura híbrida SDN/NFV que permite tener en cuenta la posición de controladores y de orquestadores en la red. Dado que los criterios de selección de ambos elementos no son los mismos, se plantean diferentes soluciones aplicables según el caso. El primer método selecciona las ubicaciones con el menor retardo entre los nodos especiales y los nodos normales. El segundo método determina las localizaciones que proporcionan el menor retardo entre los nodos normales. Por último, el tercero proporciona un retardo equilibrado entre todos los nodos, ya sean especiales o normales. Estos tres

métodos son comparados con otro que escoge la ubicación de los nodos especiales de forma aleatoria.

El resto del artículo se estructura de la siguiente manera. En la sección II se presenta el modelo de red utilizado, en la sección III el algoritmo propuesto, en la sección IV los resultados de la evaluación de los métodos que forman el algoritmo, y en la sección V el artículo finaliza con las conclusiones.

II. MODELO DE RED

La investigación se inicia a partir de una red formada por un conjunto de nodos N conectados a través de enlaces $e \in E$, modelada mediante un grafo no dirigido $G(N, E)$ cuyos enlaces llevan asociado un coste. Estos enlaces son los encargados de llevar la información de un nodo a otro mediante la ruta más corta (Método 1). En este método, $|E(n, m)|$ es el número de enlaces $e \in E$ de la ruta que une el nodo $n \in N$ con el nodo $m \in N$. A modo de aproximación se ha considerado que el retardo entre dos nodos es básicamente igual al retardo de propagación. Las rutas más cortas se han calculado a partir de estos retardos de propagación, sin carga en la red ni utilización de los enlaces. En el caso de redes WAN la mejor ubicación depende de esta latencia, la cual está fijada por la topología física. En nuestro caso, en el que analizamos redes MAN, esta consideración es una aproximación pero nos permite hallar de manera sencilla una solución inicial de ubicación de los nodos especiales. El algoritmo lleva a cabo la localización de estos nodos a partir de las siguientes características de los nodos y enlaces.

A. Características de un enlace $e(a, b) \in E$

- $t(e)$ es el tiempo de propagación de la información entre sus dos nodos $a, b \in N$.
- $|R(e)|$ es el número de rutas que atraviesan el enlace para conectar dos nodos cualesquiera de la red en el menor tiempo posible (implementado en el Método 2).

B. Características de un nodo $n \in N$

- $|E(n)|$ es el número de enlaces conectados al nodo.
- $|R(n)|$ es el número de rutas que atraviesan el nodo para conectar dos nodos cualesquiera de la red en el menor tiempo posible. Se calcula como la suma del número de rutas de cada enlace conectado al nodo (Eq. 1 y su implementación en el Método 3).

$$|R(n)| = \sum_{k=1}^{|E(n)|} |R(e_k(n))|, \forall e_k(n) \in E(n) \quad (1)$$

- $t(n, m)$ es el menor tiempo posible que tarda en propagarse la información desde el nodo n hasta el nodo m a través de los enlaces de su ruta $E(n, m)$ (Eq. 2 y su implementación en el Método 4).

$$t(n, m) = \sum_{k=1}^{|E(n, m)|} t(e_k), \forall e_k \in E(n, m) \quad (2)$$

- $t_e(n)$ es el tiempo medio que tarda en propagarse la información desde el nodo n hasta cualquier otro nodo de la red (Eq. 3 y su implementación en el Método 5).

$$t_e(n) = \frac{1}{|N|} \sum_{k=1}^{|N|} t(n, m_k), \forall m_k \in N \quad (3)$$

- $t_r(n)$ es el tiempo medio que tarda en propagarse la información desde cualquier otro nodo de la red hasta el nodo n (Eq. 4).

$$t_r(n) = \frac{1}{|N|} \sum_{k=1}^{|N|} t(m_k, n), \forall m_k \in N \quad (4)$$

- $b(n)$ es un valor que relaciona el número de rutas que pasan por él junto con su tiempo medio, considerándolo como emisor y receptor de información. Mediante el cálculo de este valor se pretende identificar a los nodos que presenten a la vez: bajo tiempo medio desde el nodo a los demás, bajo tiempo medio desde los demás nodos hasta el nodo, y menor número de rutas que pasan por él (Eq. 5).

$$b(n) = |R(n)| * t_e(n) * t_r(n) \quad (5)$$

Para conseguir un equilibrio entre ambos tiempos los igualamos. Si se cumple $t(n, m) = t(m, n)$, entonces $t_e(n) = t_r(n)$ (Eq. 6 y su implementación en el Método 6).

$$b(n) = |R(n)| * t_e(n) * t_e(n) \quad (6)$$

- $random(n)$ es una variable aleatoria asignada al nodo n .

A continuación se detalla el algoritmo que utiliza las características señaladas y que sirve para seleccionar la ubicación óptima de los nodos especiales.

III. ALGORITMO PROPUESTO

El proceso para llevar a cabo esta selección tiene cuatro fases: 1) cálculo de la ruta más corta entre cada par de nodos de la red, 2) cálculo de las características de nodos y enlaces, 3) ordenación de nodos según el resultado deseado, y 4) selección de nodos especiales.

A. Cálculo de los enlaces de la ruta más corta

En el (Método 1) se muestra el establecimiento de los enlaces de la ruta más corta entre dos nodos de la red. Estos enlaces son necesarios para calcular las características de los nodos y enlaces del modelo propuesto.

Método 1 Cálculo de $E(n, m)$

- 1: **para** cada $n \in N$ **hacer**
 - 2: **para** cada $m \in N$ **hacer**
 - 3: $E(n, m) \leftarrow EnlacesRutaMasCorta(n, m)$
 - 4: **fin para**
 - 5: **fin para**
-

B. Cálculo de las características de nodos y enlaces

Este apartado muestra la implementación de las características de los nodos y enlaces del modelo propuesto.

Método 2 Cálculo de $|R(e)|$

- 1: $|R(e)| \leftarrow 0, \forall e \in E$
 - 2: **para** cada $n \in N$ **hacer**
 - 3: **para** cada $m \in N$ **hacer**
 - 4: **para** cada $e \in E(n, m)$ **hacer**
 - 5: $|R(e)| \leftarrow |R(e)| + 1$
 - 6: **fin para**
 - 7: **fin para**
 - 8: **fin para**
-

Método 3 Cálculo de $|R(n)|$

- 1: $|R(n)| \leftarrow 0, \forall n \in N$
 - 2: **para** cada $n \in N$ **hacer**
 - 3: **para** cada $e(n) \in E(n)$ **hacer**
 - 4: $|R(n)| \leftarrow |R(n)| + |R(e(n))|$
 - 5: **fin para**
 - 6: **fin para**
-

Método 4 Cálculo de $t(n, m)$

- 1: $t(n, m) \leftarrow 0, \forall n, m \in N$
 - 2: **para** cada $n \in N$ **hacer**
 - 3: **para** cada $m \in N$ **hacer**
 - 4: **para** cada $e \in E(n, m)$ **hacer**
 - 5: $t(n, m) \leftarrow t(n, m) + t(e)$
 - 6: **fin para**
 - 7: **fin para**
 - 8: **fin para**
-

Método 5 Cálculo de $t_e(n)$

- 1: $t_e(n) \leftarrow 0, \forall n \in N$
 - 2: **para** cada $n \in N$ **hacer**
 - 3: **para** cada $m \in N$ **hacer**
 - 4: $t_e(n) \leftarrow t_e(n) + t(n, m)$
 - 5: **fin para**
 - 6: $t_e(n) \leftarrow t_e(n) / |N|$
 - 7: **fin para**
-

Método 6 Cálculo de $b(n)$

- 1: $b(n) \leftarrow 0, \forall n \in N$
 - 2: **para** cada $n \in N$ **hacer**
 - 3: $b(n) \leftarrow |R(n)| * t_e(n) * t_e(n)$
 - 4: **fin para**
-

C. Ordenación de nodos según el resultado deseado

En esta tercera fase ordenamos los nodos según una determinada característica en sentido creciente o decreciente. La elección de la característica así como su sentido de ordenamiento dará como resultado unos valores u otros. El

resultado que se desea obtener es el primer elemento a tener en cuenta a la hora de seleccionar la ubicación de los nodos. En este apartado presentamos cuatro objetivos distintos y su implementación:

1) *Bajo retardo entre nodos especiales y normales*: La característica seleccionada de ordenamiento para llevar a cabo este objetivo es el tiempo medio que tarda en propagarse la información desde el nodo n hasta cualquier otro nodo de la red, $t_e(n)$, en sentido creciente (Método 7). Por tanto, prioriza los nodos con menor retardo medio. Este método es denominado *Especial* en la sección de evaluación de los métodos.

Método 7 *Bajo retardo entre nodos especiales y normales*

```

1: para  $m = 1 \rightarrow |N| - 1$  hacer
2:   para  $n = m + 1 \rightarrow |N|$  hacer
3:     si  $t_e(N[m]) > t_e(N[n])$  entonces
4:        $aux \leftarrow N[m]$ 
5:        $N[m] \leftarrow N[n]$ 
6:        $N[n] \leftarrow aux$ 
7:     fin si
8:   fin para
9: fin para

```

2) *Bajo retardo entre nodos normales*: La característica seleccionada también es $t_e(n)$ pero ordenada en sentido decreciente (Método 8). Por lo tanto, prioriza los nodos con mayor retardo medio. Este método es denominado *Normal* en la sección de evaluación de los métodos.

Método 8 *Bajo retardo entre nodos normales*

```

1: para  $m = 1 \rightarrow |N| - 1$  hacer
2:   para  $n = m + 1 \rightarrow |N|$  hacer
3:     si  $t_e(N[m]) < t_e(N[n])$  entonces
4:        $aux \leftarrow N[m]$ 
5:        $N[m] \leftarrow N[n]$ 
6:        $N[n] \leftarrow aux$ 
7:     fin si
8:   fin para
9: fin para

```

3) *Retardo balanceado entre todos los nodos*: La característica de ordenamiento seleccionada para llevar a cabo este objetivo es el valor que relaciona el número de rutas que pasan por él junto con su tiempo medio, considerándolo como emisor y receptor de información, $b(n)$, en sentido decreciente (Método 9). Por tanto, prioriza los nodos con menor valor. Este método es denominado *Balanceado* en la sección de evaluación de los métodos.

4) *Orden aleatorio*: Su característica es un valor aleatorio asignado al nodo, $random(n)$, y en sentido decreciente (Método 10). Este método es denominado *Aleatorio* en la sección de evaluación de los métodos.

D. Selección de nodos especiales

Una vez que tenemos la lista ordenada según el método deseado, escogemos sus k primeros elementos, siendo k el número de nodos especiales y $k - |N|$ el de nodos normales.

Método 9 *Retardo balanceado entre todos los nodos*

```

1: para  $m = 1 \rightarrow |N| - 1$  hacer
2:   para  $n = m + 1 \rightarrow |N|$  hacer
3:     si  $b(N[m]) > b(N[n])$  entonces
4:        $aux \leftarrow N[m]$ 
5:        $N[m] \leftarrow N[n]$ 
6:        $N[n] \leftarrow aux$ 
7:     fin si
8:   fin para
9: fin para

```

Método 10 *Orden aleatorio*

```

1: para  $m = 1 \rightarrow |N| - 1$  hacer
2:   para  $n = m + 1 \rightarrow |N|$  hacer
3:     si  $random(N[m]) > random(N[n])$  entonces
4:        $aux \leftarrow N[m]$ 
5:        $N[m] \leftarrow N[n]$ 
6:        $N[n] \leftarrow aux$ 
7:     fin si
8:   fin para
9: fin para

```

IV. EVALUACIÓN

La ejecuciones realizadas mediante el simulador [8] nos muestran el rendimiento de los métodos presentados. Esta herramienta, desarrollada por los autores de este artículo, ha sido adaptada a los elementos y situaciones que son objeto de estudio en este artículo. La herramienta ha sido programada en Java, un lenguaje de programación orientado a objetos (POO), siendo ésta la política de diseño de los elementos que componen la red.

Para evaluar los métodos del algoritmo se han utilizado redes de 20 nodos cuyo grado varía entre 4 y 10. La distancia de cada enlace se selecciona de forma aleatoria entre 1 y 50 Kms. Nuestra intención es modelar redes de área metropolitana de aproximadamente unos 20 nodos. Estas redes permiten alcanzar un diámetro en torno a los 50 km. Este diámetro se considera suficiente para abarcar un área metropolitana. Abarcan una ciudad y se pueden conectar varias entre sí, formando más redes. La evaluación se realiza dos veces, para 3 y 6 nodos especiales, respectivamente. Se ha establecido que los nodos especiales no puedan ser seleccionados para buscar la ruta más corta entre dos nodos normales o, entre otro nodo especial y otro normal. Los resultados mostrados son la media de los obtenidos con 1000 redes generadas aleatoriamente para cada grado de red. Se han calculado los intervalos del 95% de confianza basados en la distribución t de Student pero no se muestran para dar mayor claridad a las gráficas.

Las métricas que se analizan son: i) Retardo medio entre un nodo especial y otro normal, R_{en} , ii) Retardo medio entre dos nodos normales, R_{nn} , y iii) Valor absoluto de la diferencia de ambos retardos medios, $|R_{en} - R_{nn}|$. Acompañando a las gráficas se muestra el retardo medio entre dos nodos cualesquiera de la red, que se identifica como *Media*.

A. Retardo entre un nodo especial y otro normal

La Fig. 3 muestra que el método *Especial* consigue los menores retardos mientras que el *Normal*, su opuesto, obtiene

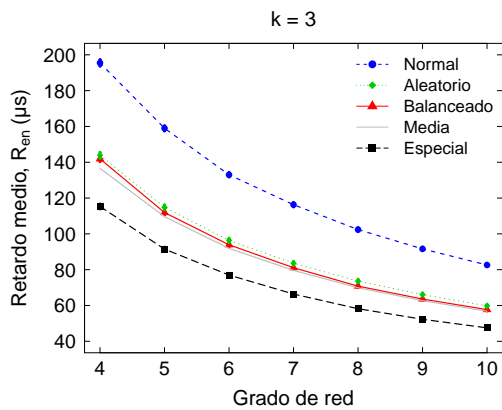


Fig. 3. Retardo medio entre un nodo especial y otro normal, con $k = 3$ nodos especiales y 17 normales.

los mayores. Por ejemplo, la diferencia de ambos valores es $80 \mu s$ cuando el número medio de enlaces conectados a cada nodo es 4. Esto significa que la información enviada desde un nodo especial establecido por el método *Normal* hasta un nodo normal, tiene que recorrer aproximadamente 24 km más en comparación con el método *Especial*. El método *Balanceado* muestra valores muy cercanos a la *Media*, seguidos por el *Aleatorio*.

La Fig. 4, con 6 nodos seleccionados, muestra el mismo comportamiento que en el caso de 3 nodos. Las principales diferencias observadas son las siguientes. Los retardos del *Especial* se incrementan al aumentar el número de nodos especiales. Por otra parte, los retardos del método *Normal* se reducen al disminuir el número de nodos especiales, mientras que tanto el *Balanceado* como el *Aleatorio* se alejan ligeramente de la *Media*.

B. Retardo entre dos nodos normales

La Fig. 5 muestra que el método *Normal* obtiene el menor retardo mientras que el *Especial*, su opuesto, da los mayores retardos. Sucede lo contrario que en el apartado anterior. El método *Balanceado* muestra valores prácticamente similares a la *Media*, seguido a más distancia por el *Aleatorio*.

La Fig. 6 muestra que el método *Normal* ofrece el menor retardo mientras que el *Especial*, su opuesto, obtiene los

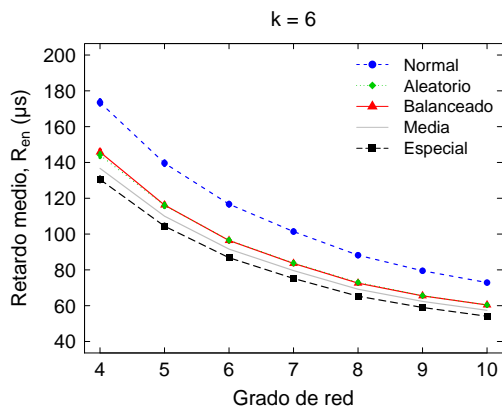


Fig. 4. Retardo medio entre un nodo especial y otro normal, con $k = 6$ nodos especiales y 14 normales.

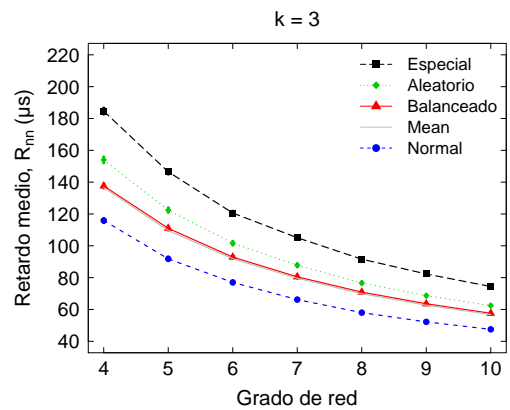


Fig. 5. Retardo medio entre dos nodos normales, con $k = 3$ nodos especiales y 17 normales.

mayores retardos. Sucede lo contrario que en el apartado anterior. El método *Balanceado* muestra valores prácticamente similares a la *Media*, seguido a más distancia por los valores de retardo del *Aleatorio*.

Comparando la Fig. 3 con la Fig. 5 (para 3 nodos especiales) o bien la Fig. 4 con la Fig. 6 (para 6 nodos especiales), se observa que en el caso de que se opte por minimizar el retardo en el plano de control (*Especial*), los caminos del plano de datos van a ser más largos provocando mayor retardo en la comunicación del tráfico de usuario.

C. Valor absoluto de la diferencia de ambos retardos

La Fig. 7 muestra la diferencia entre el retardo R_{en} y R_{nn} para el caso de 3 nodos especiales. Esta figura muestra que el método *Balanceado* ofrece la menor diferencia de retardo mientras que el *Normal*, da los mayores retardos, seguido de *Especial*. Esta sería la solución a utilizar si se desea un equilibrio entre los retardos del plano de control y los retardos del plano de datos. En este caso, los valores del *Aleatorio* se muestran claramente a distancia del *Balanceado*. Además, el valor medio deseado es 0.

La Fig. 8 muestra la diferencia entre el retardo R_{en} y R_{nn} para el caso de 6 nodos especiales. En este caso se muestra que el método *Balanceado* ofrece la menor diferencia de retardo mientras que los resultados del *Aleatorio* se alejan de los

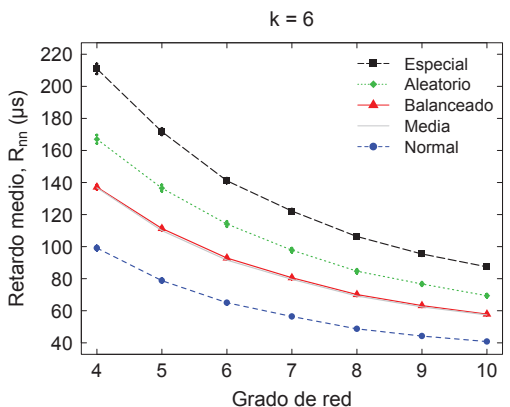


Fig. 6. Retardo medio entre dos nodos normales, con $k = 6$ nodos especiales y 14 normales.

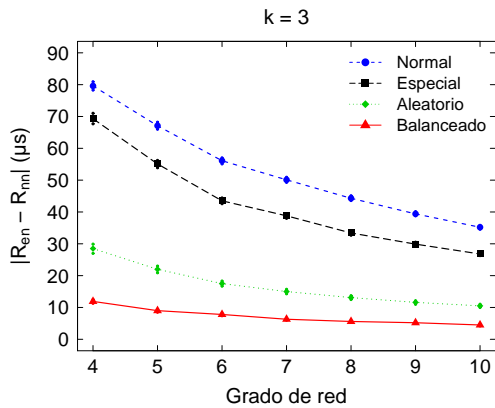


Fig. 7. Valor absoluto de la diferencia de ambos retardos medios, con $k = 3$ nodos especiales y 17 normales.

valores del *Balanceado*. En esta medición, el método *Especial* es el que muestra valores mayores seguidos del *Normal*.

V. CONCLUSIONES

En este trabajo se ha presentado una metodología para seleccionar la localización de los nodos de control de una arquitectura de red SDN/NFV, sin limitarse sólo a la localización de los controladores SDN sino extendiendo la problemática a la ubicación de los orquestadores NFV. Se ha creado un algoritmo que implementa el cálculo de las características de nodos y enlaces, con cuatro métodos de ordenación para la selección de los nodos dedicados a controlar la red (nodos especiales), según un determinado objetivo para cada uno de ellos: bajo retardo entre nodos especiales y normales; bajo retardo entre nodos normales; retardo balanceado entre todos los nodos; y ordenamiento aleatorio. Las métricas analizadas para estos métodos son: el retardo medio entre un nodo especial y otro normal, el retardo medio entre dos nodos normales y el valor absoluto de la diferencia de ambos retardos medios. La comparación de resultados refleja el objetivo de cada método, mostrando así la importancia de los criterios tenidos en cuenta a la hora de seleccionar la ubicación de los nodos especiales de la red.

La metodología creada para seleccionar la localización más adecuada de los nodos especiales y normales de una arqui-

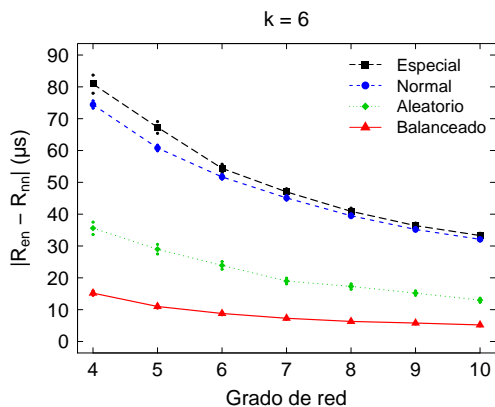


Fig. 8. Valor absoluto de la diferencia de ambos retardos medios, con $k = 6$ nodos especiales y 14 normales.

tectura mixta SDN/NFV, ha permitido extraer la conclusión que el método *Balanceado* ofrece un mayor equilibrio en el retardo entre nodos especiales y generales. Esta solución asegura que todos los nodos de la red están gestionados o controlados por un nodo especial y al mismo tiempo el retardo en el plano de control y en el plano de datos difieren mínimamente. Además, esta diferencia se mantiene independiente del grado medio de la red.

El proceso comparativo con respecto al método *Aleatorio* ayuda a comprender la necesidad de elegir los nodos bajo un criterio específico, sobre todo cuando aumenta el número de nodos especiales en la red.

En este artículo se presentan dos métodos cuyos propósitos son mutuamente excluyentes, *Especial* y *Normal*, junto con un tercero, *Balanceado*, cuyo propósito es encontrar un equilibrio entre los dos primeros. Se puede considerar que un orquestador NFV y un controlador SDN no realizan el mismo tipo de funciones ni tienen los mismos requisitos, por lo que el diseñador de la red es el encargado de seleccionar el método apropiado. Por ejemplo, para seleccionar el orquestador NFV se puede utilizar el método *Especial*, y una vez éste ha sido seleccionado, buscar los controladores con el método *Balanceado*.

Como trabajo futuro se propone incrementar la complejidad del algoritmo añadiendo una o varias restricciones de retardo como parámetro de diseño. Asimismo, se propone añadir en el algoritmo el uso de los recursos de la red por parte de los usuarios para analizar cómo influye el estado de la red en el retardo y, por lo tanto, en la ubicación de los nodos de control.

AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad del Gobierno de España a través del proyecto TEC2013-47960-C4-1-P.

REFERENCIAS

- [1] Nunes, B.A.A. ; Mendonca, M. ; Xuan-Nam Nguyen ; Obraczka, K. ; Turletti, T, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," *Communications Surveys & Tutorials, IEEE*, vol. 16 , no. 3, pp. 1617-1634, 2014.
- [2] ETSI NFV specifications, <http://community.comsoc.org/blogs/alanweissberger/etsi-network-functions-virtualisation-nfv-completes-first-phase-work-while-ope>, accessed on 17 May 2015.
- [3] Corporation Intel. "Open, simplified networking based on sdn and network functions virtualization," *White Paper from Intel Architecture Processors, Networking and Communications*, 2013.
- [4] B. Heller, R. Sherwood, N. McKeown, "The controller placement problem," *Proceedings of the First Workshop on Hot Topics in Software Defined Networks, HotSDN12*, ACM, New York, NY, USA, 2012, pp. 7-12.
- [5] Guang Yao, Jun Bi, Yuliang Li, and Luyi Guo, "On the Capacitated Controller Placement Problem in Software Defined Networks," *IEEE Communications Letters*, vol. 18, no. 8, August 2014.
- [6] Lange, S.; Gebert, S.; Zinner, T.; Tran-Gia, P.; Hock, D.; Jarschel, M.; Hoffmann, M., "Heuristic Approaches to the Controller Placement Problem in Large Scale SDN Networks", *IEEE Transactions on Network and Service Management*, vol. 12, no. 1, pp. 4-17, 2015.
- [7] Muller, L.F.; Oliveira, R.R.; Luizelli, M.C.; Gaspary, L.P.; Barcellos, M.P., "Survivor: An enhanced controller placement strategy for improving SDN survivability", *IEEE Global Communications Conference (GLOBECOM)*, 2014.
- [8] Aurelio Javier García, Cristina Cervelló-Pastor, and Yury Jiménez, "A modular simulation tool of an orchestrator for allocating virtual resources in SDN," *Journal of Communication and Computer*, 11(1):88-99, January 2014. ISSN 1548-7709.

Transmisión de vídeo H.264/AVC con control de tasa adaptativo en el servidor

J. Mata, E. Solórzano, Jose L. Muñoz, Juan Hernandez, Juanjo Alins, Oscar Esparza

Dept. Ingeniería Telemática, Universitat Politècnica de Catalunya

Abstract—In this paper we describe a method for transmitting H.264/AVC encoded video. The data transfer is performed by adapting data transmission to network performance in order to ensure continuous transmission. The data transfer between server and client terminals is performed by TCP sockets. The original video is encoded in H.264/AVC for different levels of bitrate using the GStreamer library. Then each encoded video is segmented at GOP level. The purpose of segmenting the video is to facilitate switching between different video qualities to adapt the bitrate to the variable network capacity, and using a control in the server side. Segmenting the encoded video has the advantages of scaling the digital video service and the maximum use of network resources.

Index Terms—Digital video broadcasting, H264/AVC, Adaptive Bitrate Streaming

I. INTRODUCCIÓN

En Internet el vídeo streaming puede presentar cortes o esperas durante su reproducción. Las causas pueden ser diversas, por ejemplo, lentitud y variabilidad en la velocidad de conexión, poca capacidad en el servidor, que el recurso sea muy demandado, o que el terminal del usuario tenga poca capacidad, entre otras [1][2].

En estos momentos las compañías proveedoras de vídeo streaming cuentan con servidores con capacidad para atender gran demanda de vídeos. Además, en los últimos años la capacidad de los terminales de los usuarios finales se ha incrementado considerablemente. Todas estas mejoras han favorecido al servicio de vídeo streaming, quedando como principal causa de problemas en la reproducción, la variabilidad de la velocidad de conexión.

En este artículo se propone un método para transmitir vídeo adaptando la transferencia de datos a la variación de la velocidad de conexión [3]. Este nuevo método es una nueva variante de la técnica Adaptive Bitrate Streaming. Los resultados concretos de esta investigación son un conjunto de herramientas y librerías para procesar streams H.264/AVC.

II. MÉTODOS IMPLEMENTADOS BASADOS EN ADAPTIVE BITRATE STREAMING

El Adaptive Bitrate Streaming representa un avance

fundamental respecto a la clásica descarga progresiva de streaming. Gracias a Adaptive Bitrate, la fuente de vídeo se adapta sobre la marcha, con lo que el usuario puede experimentar la calidad máxima que permite el ancho de banda [4][5].

Existen varias implementaciones, entre las más importantes MPEG-DASH, adobe Dynamic Streaming for Flash, Apple HTTP adaptive Streaming para iOS, y Microsoft Smooth Streaming. En los siguientes apartados se indican las principales características de cada uno de ellos.

A MPEG-DASH

DASH ofrece una solución estándar para la transmisión eficiente y fácil de multimedia utilizando la infraestructura HTTP disponible, especialmente servidores y CDN (Content Delivery Networks), además de proxies, cachés, etc. Permite el despliegue de los servicios de streaming utilizando la infraestructura de Internet existente sin disposiciones especiales. Es compatible tanto para la transmisión bajo demanda, como en vivo. Dado que DASH es un formato y codec general, y se puede utilizar con cualquier formato de medios de comunicación, cuenta con disposiciones específicas para el formato MPEG-4 y MPEG-2. [6][7][8]

B Adobe HTTP Dynamic Streaming

HTTP Dynamic Streaming es el proceso de transmisión de vídeo que cambia dinámicamente entre diferentes flujos de diferente calidad y tamaño durante la reproducción. El objetivo es adaptar la reproducción al ancho de banda disponible y a la capacidad del dispositivo [9]. Las últimas versiones de Flash Player y Flash Media Server admiten el streaming adaptativo a través del protocolo RTMP tradicional, así como HTTP, de forma similar a las soluciones basadas en HTTP de Apple y Microsoft [10].

C Apple HTTP Live Streaming

HTTP Live Streaming se compone de tres partes: el componente de servidor, el componente de distribución, y el software de cliente. El componente de servidor se encarga de tomar los flujos de entrada y codificarlos digitalmente. Luego los encapsula en un formato adecuado para la entrega, y finalmente prepara la encapsulación para su distribución. El

componente de distribución se compone de servidores web estándar, que son responsables de aceptar las solicitudes de cliente y la entrega de medios preparados. Para la distribución a gran escala, también se pueden utilizar las redes de última generación u otras redes de entrega de contenido. El software cliente es responsable de determinar los medios adecuados para solicitar, la descarga de esos recursos, y el posterior montaje de los mismos de modo que los medios de comunicación pueden ser presentados al usuario en un flujo continuo. El software de cliente se incluye en iOS 3.0 y posteriores y las computadoras con Safari 4.0 o posterior. [11]

D Microsoft Smooth Streaming

Smooth Streaming es una extensión de IIS Media Services que permite el streaming adaptativo de los medios de comunicación a los clientes a través de HTTP [12]. La especificación de formato se basa en el formato de archivo ISO de medios de base y estandarizado por Microsoft como el Interoperable Format de archivo protegido [13][14].

III. H.264/AVC

La tasa de un vídeo codificado depende de la complejidad de los fotogramas del vídeo original y el parámetro de cuantificación (QP) usado por el codificador H.264/AVC. En la Fig. 1 se muestra la característica bitrate-QP y el efecto de la complejidad del vídeo a codificar. Para un valor de QP la variación de la tasa depende de la complejidad de la fuente. Para un valor pequeño de QP el rango de bitrates es alto y viceversa, para un valor grande de QP el rango de bitrates del vídeo codificado es bajo. El VCL (Video Coding Layer) de H.264/AVC especifica que el incremento porcentual del parámetro de cuantificación (QP) implica un decremento porcentual similar de la tasa del vídeo codificado, aunque en última instancia dependerá del codificador y de los parámetros elegidos en sus algoritmos.

La especificación del NAL (Network Abstraction Layer) [15] de H.264/AVC indica que la unidad básica de transporte es la unidad NAL. Las unidades NAL se pueden clasificar en dos grupos: unidades NAL con contenido VCL (NAL VCL) y unidades NAL sin contenido VCL (NAL nonVCL). Dependiendo de las necesidades de transporte la estructura del stream de vídeo codificado puede cambiar, es decir, usando las unidades NAL nonVCL se pueden crear cápsulas de vídeo independientes (con al menos una unidad NAL IDR, sus respectivas unidades NAL SPS y PPS y unidades NAL CS con slices P y B). Estas cápsulas se denominan grupos de imágenes (Group Of Pictures – GOP).

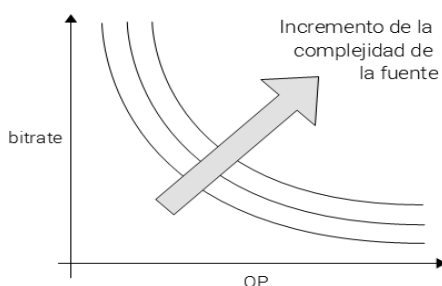


Fig. 1: Curvas bitrate-QP

IV. CONTROL DE TASA EN EL SERVIDOR

Para implementar nuestro método basado en la técnica Adaptive Bitrate Streaming necesitaremos controlar la tasa del vídeo a transmitir. Ya sabemos que con el parámetro de cuantificación (QP) podemos modificar el rango de bitrates del vídeo codificado. Además debemos ser capaces de cambiar dicha tasa dinámicamente durante la transmisión sin que el usuario perciba cortes. Por ejemplo, si de una capacidad de 2 Mbps pasamos a 100 Kbps, debemos bajar la tasa de transmisión de vídeo para que pueda ser transmitido sin detener su reproducción. Para implementar este cambio dinámico de tasa durante la transmisión planteamos usar varias codificaciones distintas de la misma secuencia de vídeo, cada una con un QP diferente.

Los actuales métodos que implementa la técnica Adaptive Bitrate Streaming tienen en común la ubicación del control de tasa en el cliente. En cambio, nuestra propuesta ubica dicho control en el servidor. De este modo, será el servidor el responsable de realizar las medidas de ancho de banda y de adaptar la transmisión del vídeo a las mismas. Con ellos, los terminales quedan liberados de tener que conocer cuáles son los medios audiovisuales disponibles en el servidor para su reproducción.

Dado que la medición de la tasa de transmisión es la misma tanto si se mide en el servidor como en el cliente, situar dicho control en el servidor presenta varias ventajas estratégicas. Por ejemplo, el despliegue de esta tecnología implicaría cambios en el software de reproducción del cliente. Obviamente, esto supondría una carga adicional sobre el servidor. En todo caso, aumentar la capacidad en la parte servidor siempre es más fácil, y realizar dichos cambios simplemente está en la mano del proveedor de servicios.

A. Arquitectura de Software

La arquitectura del software para implementar el método de transmisión debe contar con lo siguiente:

- Codificador H.264/AVC de tasa variable: codifica el vídeo con un QP fijo. El vídeo codificado deberá contar con múltiples GOPs.
- Selector por tasa: elige un vídeo codificado según la tasa de transmisión.
- Medidor de tasa de transmisión: mide la tasa de transmisión para cada conjunto de datos enviados.
- Gestor de transmisión de datos: se encarga de gestionar las conexiones de los procesos que reciben los datos de vídeo codificado.
- Receptor de datos: solicita datos al gestor de transmisión de datos. Los datos recibidos los

guarda en una unidad de almacenamiento.

La arquitectura funcional se muestra en la Fig. 2. A continuación explicamos los módulos que la conforman.

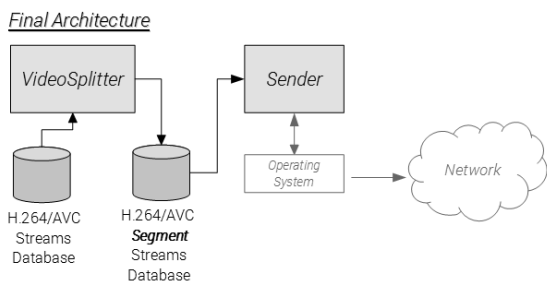


Fig. 2: Esquema de la arquitectura

1 Herramienta de División (VideoSplitter)

La exploración de los vídeos codificados y división en GOPs la realizará una única herramienta una sola vez por cada vídeo codificado antes de la transmisión. Es decir, dichos medios estarán previamente disponibles en la parte servidor.

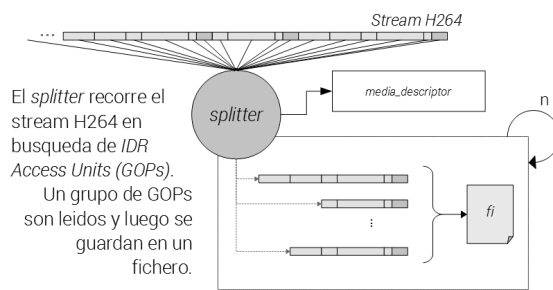
Para dividir un vídeo codificado H.264/AVC se deben de explorar e identificar los GOPs. Recordemos que un GOP es una secuencia de cuadros codificados de tipo I, P y B (p.e. IBBBPBBBBPBBBBPBBBBP) que equivale a la secuencia de unidades NAL de tipo IDR, SPS, PPS y CS.

El cambio dinámico de tasa se hará a nivel de GOP. Para pasar de la actual secuencia de GOPs de un rango de bitrates a otra secuencia de GOPs de otro rango de bitrates se puede emplear un mapa de GOPs por cada vídeo codificado, o bien un grupo de ficheros que contengan grupos de GOPs. La primera opción, el uso de un mapa, es factible pero no escalable por la limitación del número de descriptores de fichero que tienen los sistemas operativos unix-like [16] y el tiempo de uso de un fichero durante una transmisión. La segunda opción, dividir el vídeo codificado en varios ficheros más pequeños, es más escalable y es la que vamos a utilizar. Esta es la misma estrategia que usa el estándar MPEG-DASH. Con ésta división, el tiempo de uso de un fichero pequeño es menor que si el fichero contuviera todo el vídeo codificado.

En la Fig. 3 se muestra un diagrama que explica el funcionamiento de la herramienta de división de vídeo codificado. La herramienta usa un fichero de configuración que contiene: la ruta del fichero de vídeo codificado H.264/AVC a dividir, el número de GOPs que contendrá cada fichero, y la ruta donde se depositaran los ficheros que contienen los fragmentos de vídeo codificado. En nuestro caso, los ficheros son segmentos de 4 segundos, con 24 GoPs por segundo codificados con pasos de cuantificación de valor 10, 20, 30 y 40 cada uno. Como resultado, los bitrates

generados por el codificador para cada nivel de calidad de un segmento son de 2900, 2000, 400 y 100 Kbps, respectivamente. Esta elección permitirá al servicio de streaming conmutar de calidad al final de cada segmento, enviando el siguiente segmento con el nivel de calidad que más se ajusta al ancho de banda disponible en ese momento.

Después de la ejecución del VideoSplitter, en la carpeta destino se ubicarán todos los ficheros segmento del mismo contenido en una carpeta, donde residirán los ficheros codificados para cada nivel de calidad, y un fichero descriptivo (que contiene el número de ficheros generados, el prefijo que identifica a cada fichero segmento y la ruta de la carpeta que contiene a los ficheros). Este fichero será de utilidad para el proceso que gestiona el envío de vídeo codificado.



Al final del proceso splitter se habrán generado n ficheros y un fichero descriptivo.

Fig. 3: Esquema de funcionamiento de la herramienta de división de streams H.264/AVC

2 Sender

Cuando se envía un dato por un medio compartido es usual que haya pérdidas y se reenvíen datos. Estos dos eventos se pueden gestionar directamente, o bien se puede delegar dicha gestión al sistema operativo. En nuestro caso elegimos la segunda estrategia porque el trabajo desarrollado se centra en enviar vídeo codificado H.264/AVC adaptado a la capacidad del medio de transmisión. Es por eso que para la implementación usaremos sockets TCP para la petición y envío de datos, disponibles para el sistema operativo elegido para el desarrollo (linux 2.6.18).

En la Fig. 4 se muestran los bloques del proceso que gestiona el envío de datos. El servidor TCP Socket Server está a la espera de nuevas peticiones de un nuevo vídeo realizadas por un cliente. Una vez se inicia la transmisión como respuesta a una solicitud se van enviando bloques consecutivos de contenido audiovisual. La calidad de cada nuevo bloque a enviar queda determinada por la selección realizada a través del filtro que depende de los parámetros

medidos durante la transmisión del bloque anterior. Por lo tanto, durante cada bloque de datos que envía el *sender* se realiza una medición de la tasa entregada a la red (en Kbps). Para evitar la aparición de discontinuidades en el proceso de streaming, el *sender* incorpora internamente un buffer de espera entre la llegada del selector y el socket TCP establecido con el cliente. La medida realizada condicionará la decisión del selector para elegir el segmento del video codificado que se puede enviar según el desempeño actual de la red.

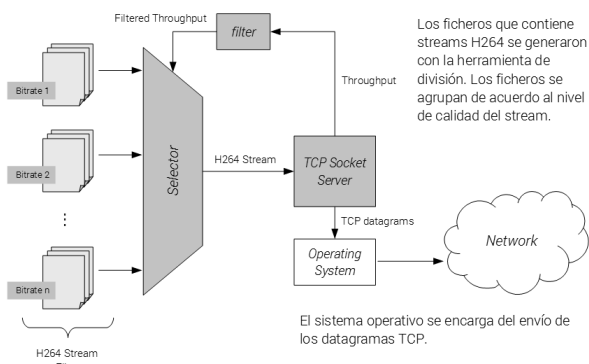


Fig. 4: Esquema del proceso de envío de datos (Sender)

Antes de lanzar el proceso se debe verificar los valores de los parámetros del fichero de configuración (la ruta de los ficheros segmento del video codificado y el puerto por donde se atenderán a los clientes TCP). Si se lanza otra instancia del proceso se deberá cambiar el puerto de escucha.

3 Receiver

En el lado del cliente, se ejecuta el proceso Receiver, encargado de establecer la conexión con el proceso de gestión de transmisión de datos que se encuentra en el terminal servidor, y de leer los datos del socket TCP y registrarlos. Al usar TCP/IP, se ha delegado al sistema operativo el control de pérdidas y petición de retransmisión. El funcionamiento es el siguiente. La aplicación lee la configuración de un fichero de inicialización que contiene la dirección IP del servidor y el puerto TCP por donde el servidor espera nuevas conexiones.

V. RESULTADOS DE LAS PRUEBAS

A Entorno de Pruebas

Para verificar el método implementado necesitaremos un entorno de pruebas que permita simular variaciones del ancho de banda de la red que comunica el terminal servidor (donde se ejecuta el proceso de envío de video codificado H.264/AVC) con el terminal cliente (donde se ejecuta el proceso de recepción de video codificado).

A día de hoy, existen varios emuladores que permiten crear redes con diversas configuraciones y dispositivos de red. Con estos emuladores se pueden realizar mediciones de tráfico para todos los protocolos de red. Además, muchos de ellos permiten la ejecución de sistemas operativos completos en los terminales y dispositivos de red. De todos los emuladores evaluados, hemos elegido Virtual Network User-Mode Linux (VNUML) [17][18] porque nos permitía modificar el comportamiento de los interfaces de los dispositivos de red, y que los terminales implementaran Linux.

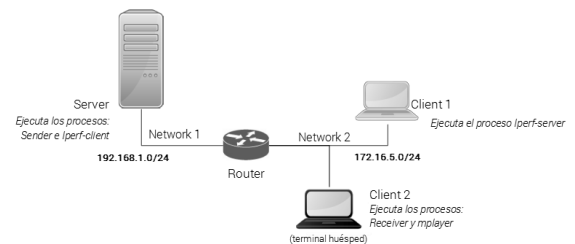


Fig. 5: Diagrama del escenario de pruebas

B Escenario: Adaptación al cambio de ancho de banda de la red del cliente

La Fig. 5 muestra el escenario de pruebas. El router dispone de un linux kernel 2.6.18, así como las herramientas para controlar el tráfico de sus interfaces de red. Nos interesa modificar el ancho de banda de la red del cliente. Para que el ancho de banda se mantenga fijo, limitaremos la velocidad a la que los paquetes son extraídos del interfaz que se conecta a la red del cliente [19][20][21].

Para observar el efecto de este escenario de pruebas sobre la transferencia, hemos probado con un video de complejidad constante cuyos bitrates son muy poco variantes en torno a sus valores medios. Los pasos de cuantificación elegidos para la codificación H.264 son: 10, 20, 30 y 40. Como resultado, los bitrates medios generados por el codificador para cada nivel de calidad de un segmento son de 2900, 2000, 400 y 100 Kbps, respectivamente. Asociado a cada paso de cuantificación fijamos un ordenación de los niveles de calidad de mejor a peor numerados del 4 al 1 en cada caso. Por tanto, el nivel de calidad que se muestra en las gráficas se corresponde inversamente con el paso de cuantificación QP. Esta es la forma habitual de representar este tipo de gráficas.

Con el fin de observar el comportamiento del servicio y sus tiempos de respuesta frente a fluctuaciones de ancho de banda se ha especificado un perfil de velocidades de la red como se muestra en la Fig. 6. Este perfil se lleva a cabo limitando la velocidad a la que los paquetes son enviados por la interfaz que se conecta a la red del cliente de forma programada.

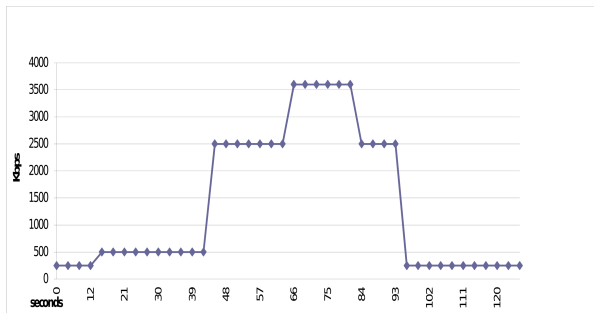


Fig. 7: Perfil de velocidades generado al modificar el ancho de banda del interfaz del router que se conecta con la red cliente.

Durante la transferencia, se puede apreciar que el algoritmo selecciona las fuentes más convenientes (ver Fig. 7). Cada vez que se realiza una medida se evalúa si la calidad es la mayor posible para la capacidad disponible. Cabe destacar que un cambio de ancho de banda no necesariamente genera un cambio inmediato en la fuente seleccionada debido al tiempo de estimación del ancho de banda disponible y a la discretización de los tiempos de conmutación de nivel de calidad impuestos en el proceso de fragmentación llevado a cabo por el *VideoSplitter*, que en nuestro caso son 4 segundos.

En la práctica, como se puede apreciar en las gráficas, las transiciones en el ancho de banda no coinciden temporalmente con los de cambio de la fuente. Estas transiciones tienen un retardo próximo a los 10 segundos. Esto se debe a que el algoritmo de selección de fuente depende de la tasa medida cuya sensibilidad depende fuertemente del tamaño del buffer de envío del sistema operativo en el servidor. El tamaño de este buffer se puede ajustar a través del uso de las opciones en la llamada al socket durante su creación. En nuestro caso, se ha mantenido el valor por defecto (4MBytes) del sistema operativo Linux. Si se pretende reducir los tiempos de respuesta a las variaciones en la red se debe ajustar coordinadamente el tamaño del buffer de escritura del socket del sistema operativo y la duración de

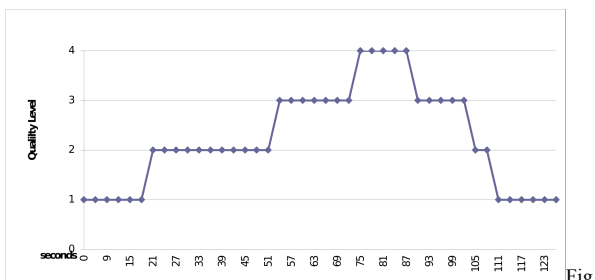


Fig. 8: Nivel de calidad del vídeo seleccionada durante la transmisión

la codificación de los segmentos.

C Escenario: Adaptación al uso compartido de la red cliente

Una de las principales características de una red es que más de un terminal esté usando el medio de manera simultánea. Al compartir la red, los terminales podrían no tener una velocidad de transmisión y recepción garantizadas. Hemos realizado pruebas para evaluar cómo se comporta nuestro método en caso de que exista tráfico interferente, y ver cómo se adapta a las variaciones de la capacidad de la red cuando

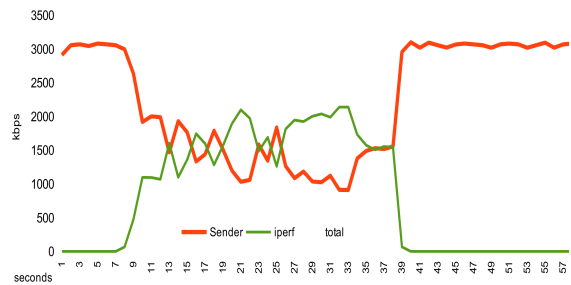


Fig. 6: Uso de la capacidad de la red por los procesos *sender* e *iperf*. Datos obtenidos con un analizador de redes.

otras conexiones TCP también la ocupan.

Para simular el uso compartido de una red y la ocupación de la mayor parte de la capacidad de la red por un terminal ejecutaremos en uno de los terminales de la red del cliente el proceso *iperf* en modo servidor (en modo TCP compartiendo la capacidad del canal) y en el terminal servidor, de la red del servidor, el proceso *iperf* en modo cliente (enviando datos durante 30 segundos). La comunicación entre los procesos *iperf* se iniciará en un periodo intermedio de la transmisión de vídeo realizado por el proceso *sender* (en un terminal de la red del servidor) al proceso *receiver* (ubicando en un terminal de la red del cliente). La capacidad de la red se fijó en 3000 Kbps en el interfaz eth1 del router (ver Fig. 5).

La Fig. 8 muestra el uso del medio medido con Wireshark, cuando el proceso *sender* y el proceso *iperf* comparten el medio. Se puede apreciar cómo ambos flujos TCP compiten por el medio, de manera que la capacidad total se reparte.

La Fig. 9. muestra la secuencia de calidades de vídeo seleccionadas. El algoritmo ha sido configurado para empezar probando con el nivel más bajo de calidad y si puede incrementar al nivel siguiente. El algoritmo incrementa la calidad si la transferencia del segmento actual se ha realizado a una velocidad superior a la necesaria para enviar el siguiente segmento de calidad superior. Cuando la capacidad del medio se ve reducida por el uso de otro proceso, el algoritmo del *sender* selecciona calidades de vídeo de menor tasa y se adapta perfectamente a las nuevas condiciones de

Estudio sobre necesidades, hábitos, preferencias y expectativas de los usuarios con respecto a la sincronización multimedia en escenarios híbridos

Fernando Boronat, Mario Montagud, Marc Martínez, Dani Marfil

Departamento de Comunicaciones,
Immersive Interactive Media (IIM) R&D Group
Campus de Gandia - Universitat Politècnica de València
{fboronat@dcom, mamontor@, marmarc3@, damarre@ei}.upv.es

Resumen- Currently, the amount and diversity of media content, together with the ubiquity and the different connectivity technologies available in the current consumption devices, enables the combination and coordination of broadcast and broadband (hybrid) networks for delivering hybrid media content. The use of Smart TVs and secondary devices (e.g., tablets, smartphones...) for that purpose is significantly growing, enabling personalized and enriched media experiences to users. Moreover, the media consumption paradigm is evolving from lonely consumer experiences to group shared experiences. Users can share media experiences while being either physically together or geographically separated.

This paper presents the preliminary results of a research study focused on collecting the users' preferences, needs, opinions and expectations about several use cases, in which hybrid and inter-destination media synchronization are needed. Such results will help to determine what needs to be developed, offered and improved regarding media synchronization, in order to provide end-users with a good QoE.

Palabras Clave- Sincronización híbrida, HbbTV, IDMS, TV Social, Televisión interactiva, Smart-TV, TV híbrida, TV conectada

I. INTRODUCCIÓN

La gran cantidad y diversidad de contenidos multimedia, unido a la amplia ubicuidad y múltiple conectividad de los dispositivos de consumo actuales, posibilitan el uso combinado y coordinado de redes broadcast y broadband para la distribución de contenidos multimedia híbridos (es decir, provenientes de redes híbridas broadcast y broadband). En concreto, el número de TVs conectadas, así como el uso de dispositivos secundarios, está cobrando protagonismo, posibilitando un gran abanico de nuevas posibilidades y nuevos modelos de negocio.

Aunque actualmente se están gestando varias iniciativas para la TV híbrida, como el estándar HbbTV¹ (Hybrid

broadcast broadband TV [1]) o la Smart TV Alliance² [2], dichas soluciones todavía no explotan al máximo el potencial que pueden ofrecer los servicios multimedia híbridos.

Además, hoy en día, estamos asistiendo a un cambio de paradigma en cuanto al consumo de contenidos multimedia. El modelo tradicional de usuarios aislados (viendo contenido en el TV u ordenador) está migrando hacia experiencias multimedia compartidas entre grupos de usuarios que están en lugares separados (como, por ejemplo, viendo contenido en el TV u ordenador, cada uno en un lugar distinto, y comunicándose entre ellos a través de medios telemáticos y/o redes sociales). Este nuevo paradigma está originado por el deseo natural del ser humano de compartir información y consumir contenido multimedia como parte de un grupo [3]. La tecnología actual ya permite que los usuarios puedan tener experiencias en grupo, aunque estén dispersos geográficamente. Sin embargo, para que estas experiencias sean satisfactorias, personalizadas e inmersivas, todavía se debe seguir investigando para mejorar las aplicaciones y la tecnología subyacente. Aún quedan muchos retos y barreras que superar.

En este artículo se presentan los resultados preliminares de un estudio de investigación que, a partir de las respuestas de los usuarios a un cuestionario on-line sobre aspectos generales y sobre 4 casos de uso representativos, trata de identificar las necesidades, hábitos, expectativas y preferencias de los usuarios cuando consumen contenidos multimedia híbridos, tanto de forma aislada como en grupo. Dicho estudio constituye una de las tareas a realizar en un proyecto de investigación sobre sincronización de contenidos híbridos (sincronización híbrida, en adelante), en el que está trabajando nuestro grupo y que está financiado, parcialmente, por el Fondo Europeo de Desarrollo Regional (FEDER) y por el Ministerio de Economía y Competitividad (MINECO), bajo el programa de apoyo a la I+D+i, con referencia TEC2013-45492-R.

¹ HbbTV consiste en una iniciativa global cuyo fin es armonizar la entrega broadcast y broadband de servicios de entretenimiento a los usuarios a través de TVs conectadas, set-top boxes y segundas pantallas (multi-screen devices). Se trata de un estándar desarrollado por compañías líderes en el mundo audiovisual para mejorar la calidad de la experiencia (QoE) del usuario, facilitando servicios innovadores e interactivos sobre redes híbridas broadcast y broadband.

² Se trata de una alianza de empresas cuya misión es la de crear un gran ecosistema productivo para el desarrollo de aplicaciones para Smart TVs que proporcione a los fabricantes, de forma rápida, un mayor número de aplicaciones, que permita a los desarrolladores crear aplicaciones que funcionen en todas las plataformas existentes y que, al mismo tiempo, proporcione a los usuarios una fuente rica de contenidos y servicios.

En dicho proyecto se van a abordar varios retos tecnológicos en el contexto de las comunicaciones multimedia híbridas, con tal de superar limitaciones importantes de las soluciones existentes. Uno de los retos principales del proyecto es la sincronización del consumo de contenidos híbridos, tanto en un mismo dispositivo (por ejemplo, una Smart TV), como en varios dispositivos (por ejemplo, un TV y una tableta), y, además, tanto en un mismo sitio (por ejemplo, en la sala de estar de una vivienda) como en diferentes sitios remotos (por ejemplo, varios usuarios que están cada uno en su vivienda). Según el caso, el tipo de sincronización necesario es distinto ([4]), abarcando la sincronización entre flujos reproducidos en un mismo dispositivo (sincronización inter-flujo o *inter-stream sync*), la sincronización entre flujos reproducidos en distintos dispositivos cercanos (sincronización entre dispositivos, *inter device sync* o IDES) y la sincronización entre flujos reproducidos en diferentes dispositivos localizados geográficamente en sitios remotos (sincronización entre destinatarios, *inter-destination media sync* o IDMS [5])

Para abordar dicho reto, se decidió realizar, en primer lugar, el estudio cuyos resultados preliminares se presentan en este artículo, con tal de constatar las necesidades, preferencias y expectativas de los usuarios en el ámbito descrito previamente.

Por un lado, existen ya numerosos estudios que muestran un aumento del uso de segundas pantallas para conectarse a redes sociales y/o navegar mientras se ve la TV, como [6] [7], e incluso comparándolo con el uso de Smart TV para realizar dichas tareas [8].

Por otro lado, también existen unos pocos trabajos sobre soluciones y técnicas de sincronización de contenidos híbridos, como [4], [9], [10] y [11]. Sin embargo, los autores no han encontrado estudio alguno que se haya centrado en el análisis de las preferencias, necesidades, opiniones y expectativas de los usuarios finales, en cuanto al consumo de contenidos híbridos, lo cual ha motivado el trabajo presentado.

El estudio pretende dar respuestas, entre otras, a las siguientes preguntas: ¿Cuáles son las preferencias de los usuarios en aplicaciones o servicios multimedia donde es necesaria la sincronización híbrida?; ¿cuáles son sus necesidades en dichas aplicaciones o servicios?; ¿cuál es su opinión sobre las soluciones de sincronización híbrida implementadas en las aplicaciones actuales?; o ¿cuáles son sus expectativas de futuro sobre dichas aplicaciones?

Los resultados que se presentan, aunque no son definitivos, vislumbran o revelan los intereses, opiniones y necesidades de los usuarios y dan respuestas a dichas preguntas. También ayudarán a los desarrolladores a determinar qué aplicaciones y servicios deberán ser proporcionados y/o mejorados en el futuro con respecto a los procesos involucrados en la cadena multimedia extremo-a-extremo, incluyendo la producción, entrega y consumo de contenidos multimedia, con el fin último de proporcionar a los usuarios una mejor calidad de experiencia (*QoE* o *Quality of Experience*).

La estructura del artículo se presenta a continuación. Los cuatro casos de uso sobre los que se ha realizado el estudio se presentan en la Sección II. En la Sección III se describe la metodología seguida en el estudio, incluyendo el diseño y las herramientas utilizadas. A continuación, en la Sección IV, se

presenta la muestra de usuarios que participaron en el estudio, hasta la fecha de recopilación de los datos presentados. Los resultados para dos de los casos de uso, de los cuatro contemplados en el estudio, se presentan en la Sección V. El artículo finaliza con una serie de conclusiones, en la Sección VI, así como señalando las limitaciones del estudio y el trabajo futuro a realizar, en la Sección VII.

II. CASOS DE USO

Para realizar el estudio se diseñó un cuestionario centrado en cuatro casos de uso o escenarios típicos, que se eligieron como representativos entre los diferentes contextos donde la sincronización híbrida entre flujos y entre destinatarios es necesaria, descritos en [4]. Se han elegido estos casos de uso por considerarlos como los más populares, posiblemente, en la actualidad, tal y como se ha podido corroborar con las respuestas de los usuarios sobre su experiencia previa en cada uno de ellos.

Los casos de uso elegidos, sobre los que se ha preguntado a los usuarios, son los siguientes:

- Caso 1: Comentarista deportivo. Consiste en un usuario que visualiza un partido de su equipo deportivo favorito en TV mientras utiliza un dispositivo secundario para escuchar los comentarios sobre el partido en un canal de radio on-line alternativo.
- Caso 2: Test de inteligencia (coeficiente intelectual). Consiste en un usuario que visualiza un concurso de TV sobre preguntas de inteligencia, para obtener el coeficiente intelectual (C.I.) de los concursantes. Asimismo, dicho usuario ejecuta una aplicación en un dispositivo secundario para comprobar su C.I. con las mismas preguntas que los concursantes en plató, de manera que también puede competir contra ellos.
- Caso 3: Multi-cámara de Fórmula 1/Moto GP (Gran Premio). Consiste en un usuario visualizando una carrera de Fórmula 1 o Moto GP en TV y que, al mismo tiempo, puede seleccionar vistas de diferentes cámaras que pueden ser visualizadas en uno o varios dispositivos secundarios.
- Caso 4: TV Social, la final del Mundial de Fútbol. Consiste en un usuario visualizando la final del mundial de fútbol en TV y que, al mismo tiempo, está interactuando mediante una herramienta de chat, ya sea texto (por ejemplo, Whatsapp) o audio/video conferencia (por ejemplo, vía Skype), sobre la evolución del partido con amigos o familiares que están en diferentes lugares remotos.

III. METODOLOGÍA

A. Diseño

Se trata de un 'estudio descriptivo correlacional u observacional' [12], ya que se recoge información de los usuarios sin previamente haber cambiado ni influido en el entorno, es decir, no se manipula nada para realizar el estudio. Se recogen evidencias de varios casos de uso reales (es decir, del mundo real, tal y como es). Además, el estudio comprende una única interacción con cada usuario mediante la recogida de información a través de un cuestionario on-line anónimo que el usuario debe rellenar una sola vez, en un instante de tiempo determinado. Por tanto, se trata de un estudio *cross-sectional* [12].

En la próxima sección se proporciona más información y un resumen de la información general sobre el tipo de usuarios que han completado el cuestionario.

B. Herramientas utilizadas

En primer lugar, se preparó la redacción y maquetación de las preguntas del cuestionario mediante un editor de textos. Se decidió incluir gráficos e imágenes para ayudar a los participantes en el estudio a identificar cada caso de uso de la forma más rápida, intuitiva y amigable posible.

El cuestionario se publicó de forma on-line¹ a través de la plataforma SurveyMonkey^{®2}, que permite crear y alojar cuestionarios on-line para realizar todo tipo de encuestas y estudios de mercado. Dicha plataforma permite, en todo momento, poder visualizar los resultados, tanto de forma resumida como de forma individual, estadísticas, etc.

Además, la plataforma permite enviar el enlace al cuestionario on-line a través de e-mail, integrarlo en una página web o blog, y compartirlo a través de redes sociales, como Facebook o Twitter. Esto último nos ha permitido dar una gran difusión del mismo a través de los canales y redes sociales de nuestra Universidad y de los contactos de los miembros de nuestro grupo de investigación (tanto redes profesionales como no profesionales).

Los resultados preliminares fueron descargados en formato Excel, y se analizaron de forma gráfica a través de dicha herramienta.

C. Recopilación de Datos

Como instrumento de obtención de datos se ha utilizado el mencionado cuestionario on-line, abierto a todos los usuarios. De esta manera, se consideró que se podría obtener un mayor número de respuestas en menos tiempo que a través de entrevistas personales a usuarios. El cuestionario ha sido anónimo y abierto, es decir, no se necesitaba ningún tipo de credenciales para acceder al mismo y contestarlo.

El cuestionario contiene un total de 55 preguntas de tipo mixto, tanto cerradas como abiertas, incluyendo preguntas de respuesta múltiple (multi-opción). Dependiendo de las respuestas, el flujo y el número de preguntas del cuestionario a contestar puede variar para cada usuario. Se ha comprobado que todas las preguntas fueron contestadas, al menos, por más de un 35% de los usuarios.

Se efectuó una distribución de invitaciones para realizar el cuestionario a través de medios telemáticos, tales como correo electrónico, WhatsApp y otras redes sociales como LinkedIn, Facebook y Twitter.

Los datos presentados en este artículo corresponden a las respuestas recogidas de un total de 756 usuarios, obtenidas durante 5 meses, aproximadamente, desde el 11 de diciembre de 2014 hasta el 4 de mayo de 2015.

IV. MUESTRA DE USUARIOS

La muestra consiste en hombres y mujeres españoles mayores de edad. En total se dispone de las respuestas de 756 participantes, de los cuales el 45% fueron hombres, el 32%

fueron mujeres y el resto no indicaron su género. La distribución por edad, dedicación, nivel de estudios y perfil se presentan en la Figura 1. De todos ellos, un 85%, aproximadamente, residía en España, y el resto residía en el extranjero (permanente o temporalmente).

En la próxima sección se presenta un resumen y análisis de los datos recopilados para esta muestra de la población.

V. RESULTADOS

A. Datos generales

Del total de usuarios, el 91% afirmó ver la TV en cualquiera de sus variantes (tradicional, on-line, IPTV, etc.).

Un 23% de los usuarios, en ese momento, no disponía de Smart TV, pero le gustaría adquirir una pronto; un 31% no tenía una Smart TV y no tenía interés en adquirir una todavía; un 24% creía que las funciones de acceso a Internet de las Smart TV aportaban un valor añadido al visionado de TV; y un 10% tenía una Smart TV, pero no usaba las funciones de acceso a Internet.

Un 30% de los usuarios indicaron que utilizaban un segundo dispositivo mientras ve la TV; un 30% indicaron que no lo utilizan, pero pensaban que los dispositivos secundarios son útiles para proporcionar una experiencia enriquecida al visionado del contenido televisivo; y un 28% indicaron que no los utilizaban porque consideraban que les causaba distracción.

Al preguntar a los usuarios si deseaban que se les proporcionara contenido extra relacionado con el que estaban visualizando en el TV, un 13% contestaron que No, pues sólo querían ver la TV tradicional; un 16% contestaron que Sí que estaban interesados en contenido extra, pero que preferían buscarlo por sí mismos; un 27% mostraron interés, pero sólo en determinados géneros de TV; y un 30% indicaron que Sí (de los cuales un 37% indicaron que prefería verlo en la misma TV, un 24% en un dispositivo secundario, y al resto no les importaba en qué dispositivo se visualizara dicho contenido).

A los que contestaron afirmativamente a la pregunta anterior, se les preguntó en qué géneros deseaban el contenido extra. El resultado se muestra en la Tabla I. Los géneros más demandados fueron las películas, las series y las noticias.

Un 30% de los usuarios disponían y utilizaban un dispositivo secundario mientras visualizaban la TV. Entre ellos, el dispositivo más utilizado fue el móvil, con un 78.6%; en segundo lugar, una tableta, con un 54.9%; a continuación un portátil, con un 39.3%; seguido de un PC, con un 1%. El sistema operativo (S.O.) más utilizado fue Android, con un 63.6%; en segundo lugar, iOS, con un 40.8%; seguido de Windows, con un 21.8%; y, finalmente, Linux, con un 0.5%. La suma de porcentajes supera el 100% ya que se trataba de preguntas de respuesta múltiple y un mismo usuario podía utilizar varios dispositivos, cada uno con diferente S.O.

¹ A fecha de redacción del artículo, el cuestionario está accesible en: <https://www.surveymonkey.com/s/sinhib>

² <https://www.surveymonkey.com>

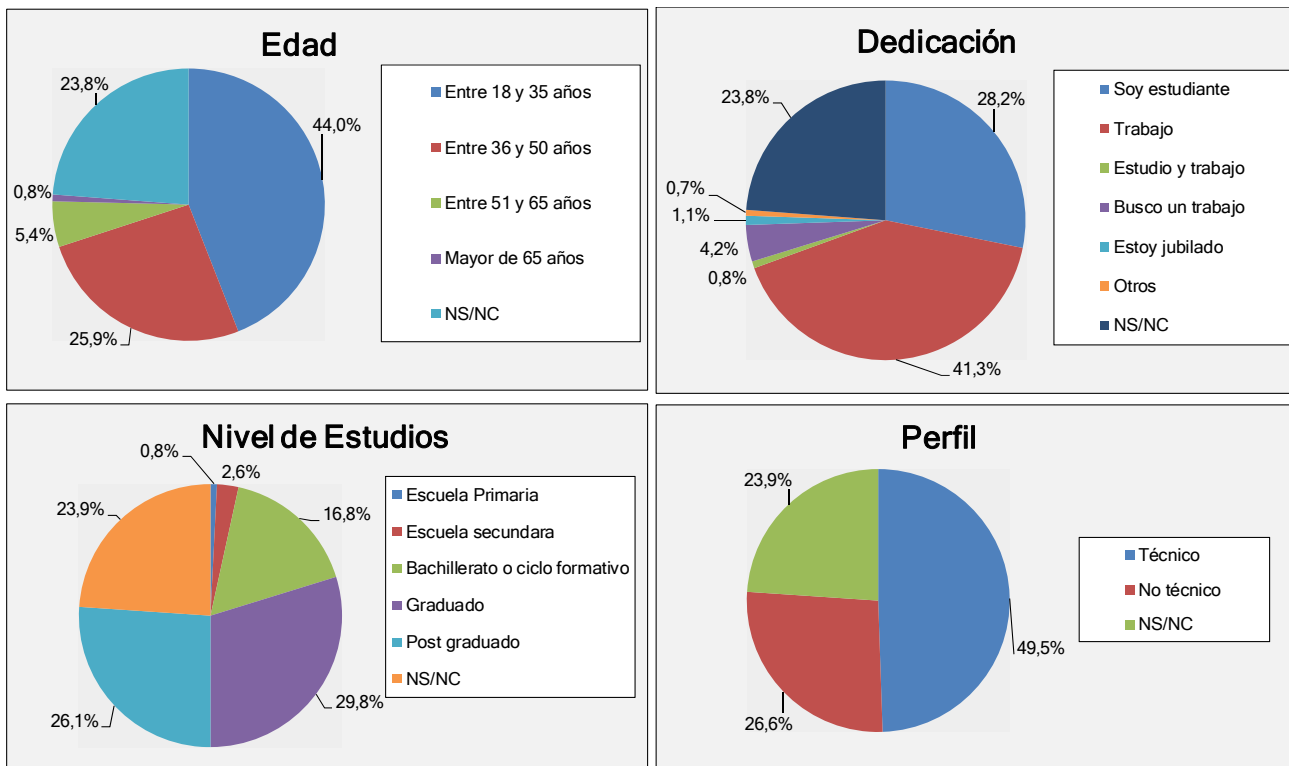


Fig. 1. Distribución de los usuarios por edad, dedicación, nivel de estudios y perfil técnico

Tabla I
GÉNEROS DEMANDADOS DE CONTENIDO EXTRA

Género	%
Reality shows	5
Películas	19.8
Noticias	15.15
Deportes	11.1
Juegos	5.5
Series	17.5
Música	12.2
No ficción	13.1
Porno	0.3

B. Expectativas de los usuarios

Se preguntó a los usuarios (pregunta de respuesta múltiple) sobre sus expectativas respecto al contenido extra relacionado, a través de una pregunta de respuesta múltiple. Las respuestas de los mismos se muestran en la Tabla II, donde se aprecia que los usuarios deseaban que, principalmente, el contenido extra estuviera actualizado, fuera fiable y estuviera sincronizado con el contenido principal.

C. Resultados de los casos de uso

En este artículo, por razones de espacio, sólo se presentan los resultados de los dos casos de uso en los que los usuarios habían participado previamente en mayor medida y, que, por tanto, evidencia su relevancia actual, que son los casos 1 y 4.

Caso 1. Comentarista deportivo

De todos los usuarios, un 10.6% indicaron que no habían tenido interés en este tipo de experiencias.

Un 24.1% de los usuarios afirmaron haber participado en experiencias similares, tales como visualizar eventos por TV (deportivos, entrega de premios Goya...) mientras escuchaban los comentarios de la radio (FM u online) sobre dicho evento. La Tabla III muestra las sensaciones experimentadas por dichos usuarios. Se aprecia que un 34.8% de los usuarios habían tenido sensaciones negativas (o muy negativas), un 38.2% las había tenido positivas (o muy positivas), y un 26.9% no se pronunciaron ni positivamente ni negativamente.

Tabla II
EXPECTATIVAS SOBRE EL CONTENIDO EXTRA

Expectativas	%
Que esté continua y automáticamente actualizado.	61.8
Que sea fiable.	49.3
Que esté sincronizado con el contenido de la pantalla principal.	47.3
Que se pueda filtrar los tipos de contenido adicional.	24.2
Que se pueda filtrar las fuentes de contenidos adicionales.	19.8
Que proporcione información complementaria	1
Nada	5.8

Tabla III
SENSACIONES PARA EL CASO 1

Sensación	%
Muy negativa	1.7
Negativa	33.1
Neutral	26.9
Positiva	33.1
Muy positiva	5.1

El motivo principal indicado por los usuarios que consideraron la experiencia como negativa fue la falta de sincronización entre la imagen de TV y el audio de la radio. Los usuarios que la consideraron como positiva alegaron que así podían elegir los comentarios de su locutor preferido o en su propio idioma y, además, un usuario señaló que ya había utilizado una herramienta para radio on-line con posibilidad de sincronización entre el audio y el vídeo (aunque no especificó cuál).

También se preguntó a los usuarios sobre los beneficios que pueden aportar experiencias como la de este caso de uso y las respuestas obtenidas se reflejan en la Tabla IV (al ser una pregunta de respuesta múltiple, la suma de los porcentajes supera el 100%). Cabe destacar que un 69.2% de los usuarios indicaron que esperaban obtener una experiencia más agradable y personalizada, un 51.1% indicaron que este tipo de aplicaciones podía contribuir a la integración social (por ejemplo: gente extranjera podía seleccionar el audio en su propio idioma, gente con problemas auditivos podían usar auriculares, etc.), y un 42.3% reflejaron que podían adquirir un mayor conocimiento sobre el evento deportivo.

Los usuarios también señalaron las limitaciones de sus experiencias en dicho caso de uso, tal y como se muestra en la Tabla V (pregunta de respuesta múltiple), entre las que destacamos la falta de sincronización (78.8%).

También se preguntó a los usuarios acerca de si creían que una tecnología adecuada para sustituir el audio por otro relacionado mientras se ve la TV podía solucionar un problema real existente. De los que contestaron a dicha pregunta, el 4% creía que dicha tecnología no es interesante en absoluto, un 29% indicaron que sí y que ya existían soluciones para ello, y un 67% indicaron que sí pero que todavía es necesario desarrollar mejores soluciones (Figura 2).

Tabla IV
BENEFICIOS DEL CASO 1

Beneficios	%
Mayor conocimiento sobre el evento deportivo.	42.3
Una experiencia más agradable y personalizada.	69.2
Puede contribuir a la integración social	51.1
Aprendizaje de idiomas	1
Poder seleccionar mejores comentaristas deportivos	0.7
No le veo beneficio alguno.	0.7

Tabla V
LIMITACIONES DEL CASO 1

Limitaciones	%
Inconvenientes técnicos, falta de sincronización.	78.8
Inconvenientes técnicos, velocidad de conexión a Internet insuficiente.	64.4
Confusión cuando se ve en TV a los comentaristas, pero se está escuchando a otros.	17.6
Puede llegar a distraer y resultar incómodo	0.3
Falta de formación tecnológica	0.3
No veo limitación alguna.	3.6

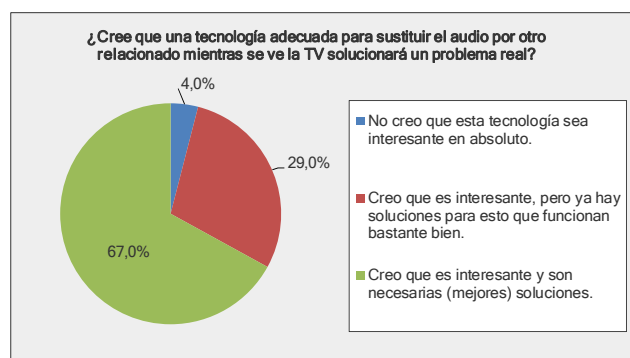


Fig. 2. Necesidad de Mejora Tecnológica para el Caso 1

Caso 4. TV Social, la final del Mundial de Futbol

Para este caso de uso, un 22% de los participantes afirmó haber tenido experiencias similares, tales como visualizar eventos deportivos, series y shows televisivos mientras interactuaba de manera remota con otras personas (mayoritariamente familiares y amigos). La Tabla VI muestra las sensaciones experimentadas por dichos usuarios.

Por un lado, el motivo principal indicado por los usuarios que consideraron la experiencia negativa (o muy negativa) fue la distracción, la falta de sincronización con los otros usuarios, la falta de fluidez del proceso de reproducción debida a los retardos, la falta de sensación de compartir una experiencia más allá de un intercambio de pensamientos, y que las comunicaciones vía texto no eran muy interactivas y preferían otras herramientas (como audio/video conferencia). Por otro lado, los usuarios que la consideraron como positiva (o muy positiva) alegaron que les proporcionó un sentimiento de estar manteniendo una experiencia compartida agradable y enriquecida con otras personas en distintos lugares y ello les mantenía 'enganchados' al programa de TV.

Para este caso de uso, sólo un 4.1% de los usuarios indicó que no tenía interés en este tipo de experiencias compartidas, sino que quería ver la TV tranquilamente de forma pasiva y aislada.

El interés mostrado por los usuarios, según el género del contenido televisivo sobre el que tener la experiencia compartida, se muestra en la Tabla VII (pregunta de respuesta múltiple). Los géneros más demandados fueron deportes, juegos, series y películas.

Como en el caso anterior, también se preguntó a los usuarios sobre los beneficios que podían aportar experiencias como la de este caso de uso y las respuestas obtenidas se reflejan en la Tabla VIII (pregunta de respuesta múltiple). Como beneficios, en este caso, el 31% de los usuarios indicó que les atraía el hecho de poder compartir experiencias y situaciones, socializar en el contexto de intereses o aficiones comunes, así como sentirse integrados como parte de un grupo, a pesar de las barreras geográficas. El 18.78% indicó que le atraía el hecho de poder compartir experiencias de TV con sus amigos o familiares de manera remota. De hecho, para un 14.15%, este era un motivo por el que ver la TV. El 12% indicó que poder 'chatear' con sus amigos mientras veían la TV les mantenía enganchados al programa que estaban visualizando. Únicamente el 2.2% opinaron que este caso de uso no les aportaba beneficio alguno.

Tabla VI
SENSACIONES PARA EL CASO 4

Sensación	%
Muy negativa	0
Negativa	8.1
Neutral	21.6
Positiva	40.5
Muy positiva	29.7

Tabla VII
GÉNEROS DEMANDADOS PARA EXPERIENCIAS COMPARTIDAS

Género	%
Reality shows	32.9
Películas	41.6
Noticias	24.3
Deportes	60.4
Juegos	52.2
Series	44.7
Música	23.1
No ficción	9
Eventos Sociales	1.2

Los usuarios también señalaron las limitaciones de sus experiencias previas en dicho caso de uso, tal y como se muestra en la Tabla IX (pregunta de respuesta múltiple).

Para este caso de uso, también se preguntó a los usuarios sobre el impacto de los retardos de las comunicaciones, así como la variabilidad de los retardos entre usuarios. De los usuarios que contestaron a la pregunta, un 9.7% afirmaron no haber percibido diferencias de retardos en este tipo de experiencias compartidas; un 38.1% indicaron que, a pesar de notar las diferencias de retardos, ello no suponía un problema serio; finalmente, un 52.3% opinaron que, según su experiencia previa, las diferencias de retardos entre usuarios son una barrera que impide disfrutar de manera satisfactoria de este tipo de experiencias multimedia compartidas.

Además, se preguntó a los usuarios qué canales de interacción les parecía que contribuyen, en mayor medida, a la “conectividad virtual” (*networked togetherness*) entre los mismos. Un 57.1% de los usuarios que contestaron a esta pregunta opinaron que los mensajes de chat de texto son el mecanismo más idóneo; mientras que un 9.8% y un 33.1% opinaron que las aplicaciones de chat de audio y de vídeo son las más apropiadas, respectivamente. En particular, WhatsApp, Facebook y Twitter son las herramientas más utilizadas con respecto a Social Media, mientras que Skype es la aplicación más utilizada para audio/video-conferencia.

También se preguntó a los usuarios sobre sus expectativas con respecto a la TV Social como herramienta que les podría proporcionar una experiencia compartida con otros usuarios remotos. Las respuestas se muestran en la Tabla X (pregunta de respuesta múltiple), donde se puede apreciar que los usuarios indicaron el deseo mayoritario de que la información se presentara a todos los que estuvieran compartiendo la experiencia de forma totalmente sincronizada. También se pudo observar en el estudio que los porcentajes, en cuanto a las preferencias de los usuarios con respecto a la herramienta de interacción a utilizar (bien integrada en la propia aplicación de TV Social o bien herramientas independientes propias, como Whatsapp o Skype), fueron muy similares.

Finalmente, se preguntó a los usuarios su opinión en cuanto al interés sobre disponer de tecnologías que

permitieran minimizar la magnitud de los retardos, así como compensar la variabilidad de los mismos entre los diferentes usuarios, para proporcionarles experiencias compartidas lo más naturales posibles. El resultado se muestra en la Figura 3.

De los usuarios que contestaron a dicha pregunta, únicamente un 4.1% creían que este tipo de tecnología no es interesante en absoluto; un 32.2% indicaron que sí es interesante y que ya existen soluciones para ello; y un 63.6% indicaron que sí es interesante y que todavía se necesita desarrollar mejores soluciones.

VI. CONCLUSIONES

En este artículo se han presentado los resultados preliminares y parciales de un estudio sobre las preferencias, necesidades, opiniones, hábitos y expectativas de los usuarios finales con respecto al consumo de contenidos híbridos, centrándose, por motivos de espacio, en dos de los cuatro casos de uso analizados y, principalmente, en aspectos relacionados con la sincronización híbrida, tanto entre flujos siendo reproducidos en mismo dispositivo como en diferentes dispositivos en un entorno local o distribuido (IDES e IDMS). Dicho estudio constituye una de las fases del proyecto de I+D financiado por FEDER y el MINECO, con referencia TEC2013-45492-R, en el que está trabajando nuestro grupo de investigación.

De los cuatro casos de uso analizados en dicho proyecto, se han obtenido y analizado las expectativas, los beneficios, las preferencias y las limitaciones detectadas por los usuarios, así como su opinión en cuanto a la necesidad de mejorar la tecnología actual para proporcionar una mejor experiencia a los usuarios. En todos ellos, tanto la sincronización de contenidos híbridos como la sincronización inter-destinatario se han identificado como retos importantes a acometer en el futuro cercano para posibilitar dichas experiencias inmersivas y personalizadas.

Tabla VIII
BENEFICIOS DEL CASO 4

Beneficios	%
Me gusta el hecho de que esta experiencia me dé una sensación de unión/compañía aunque vivamos separados.	31
Es interesante saber lo que los amigos o familiares piensan.	18.78
Tener una experiencia compartida es la razón por la que veo determinados eventos televisivos (p. ej., eventos deportivos, programas, etc.)	14.5
Chatear con mis amigos mientras veo la TV incrementa mi ‘enganche’	12
No veo ninguna limitación.	2.2

Tabla IX
LIMITACIONES DEL CASO 4

Limitaciones	%
Inconvenientes técnicos, los participantes no perciben las mismas escenas al mismo tiempo.	54
Inconvenientes técnicos, el participante A no puede acceder al mismo contenido que el participante B.	47.5
Inconvenientes técnicos, velocidad de conexión a Internet insuficiente.	52.8
Los participantes pueden tener diferentes herramientas de chat.	23.2
Distracción debido a las conversaciones.	27.9
No veo ninguna limitación.	7.3

Tabla X
EXPECTATIVAS DE LOS USUARIOS RESPECTO AL CASO 4

Expectativas	%
Obtener la misma información exactamente al mismo tiempo comparado con mis amigos/familiares.	51.7
Obtener la misma información aunque esté ligeramente desincronizada con respecto a mis amigos/familiares.	18.3
Me gustaría que se me proporcionaran herramientas de chat de texto/audio/video integradas con la experiencia compartida.	30.8
Me gustaría usar mi propia herramienta de chat, incluso sin tener una experiencia compartida sincronizada.	34.9
Nada	6.1

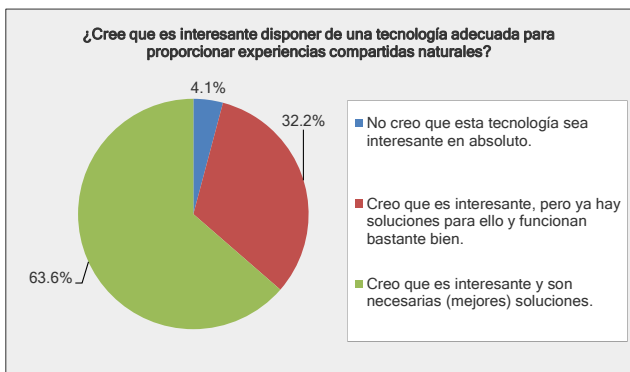


Fig. 3. Necesidad de Mejora Tecnológica para el Caso 4

De los cuatro casos de uso analizados, un 24.1% de los usuarios afirmaron haber participado en experiencias similares al caso 1; un 9.3% en experiencias similares al caso 2; un 7.8% en experiencias similares al caso 3; y un 22% en experiencias similares al caso 4. Además, muchos de ellos indicaron que su sensación al participar en ellas no fue positiva. Ello demuestra que las aplicaciones que hacen uso de contenido híbrido todavía no están implantadas de manera satisfactoria, por lo que es necesario seguir avanzando en este campo, creando contenidos, servicios y aplicaciones que proporcionen experiencias cada vez mejores a los usuarios. Aún queda mucho por mejorar en toda la cadena de distribución multimedia extremo a extremo, desde la creación de contenidos, su distribución, el desarrollo de hardware y software apropiado, etc., hasta llegar al usuario final.

La mayoría de usuarios mostró interés por los cuatro casos de uso, pero coincidieron mayoritariamente en que todavía se necesita mejorar la tecnología para ofrecer ese tipo de aplicaciones y servicios de forma satisfactoria para ellos, es decir, para poder garantizarles, al menos, una mínima QoE aceptable.

Por otro lado, cabe destacar que, aunque parezca que son las aplicaciones de ocio (relacionadas con deportes, juegos, películas, reality shows, música, etc.) las que pueden empujar en la dirección de la producción de contenidos para aplicaciones híbridas, otro tipo de géneros, como las noticias, documentales, programas de ciencia, educativos, etc., también pueden beneficiarse del acceso a contenidos complementarios obtenidos a través de Internet.

Los resultados presentados, aunque no son definitivos, ya permiten vislumbrar y revelar las necesidades, preferencias, opiniones, hábitos, demandas y expectativas de los usuarios finales en este ámbito. Estos resultados nos son muy válidos para enfocar y adaptar nuestras líneas de investigación en cuanto a sincronización multimedia híbrida se refiere, pero, además, pueden ser muy valiosos y motivadores para otros grupos de investigación relacionados, así como para todos los agentes involucrados en la cadena de distribución multimedia extremo a extremo. Asimismo, estos resultados también servirán y ayudarán a los desarrolladores a determinar qué aplicaciones y servicios deberán ser implementados, ofrecidos y/o mejorados en el futuro con respecto a los procesos de producción, entrega y consumo de contenidos multimedia para superar los desafíos identificados, con el fin último de proporcionar a los usuarios una mejor QoE.

VII. LIMITACIONES Y TRABAJO FUTURO

El estudio presentado ha sido realizado mediante un método simple, utilizando un cuestionario on-line, cuya existencia y la invitación a rellenarlo ha sido notificada a través de medios telemáticos y de redes sociales a usuarios en el ámbito de nuestra Universidad y de otras Universidades (estudiantes, profesores y personal de administración y servicios), así como en otros ámbitos cercanos a los miembros de nuestro grupo de investigación (familiares, amigos, conocidos, miembros de asociaciones comunes, etc.). Los usuarios no han sido elegidos siguiendo ningún patrón ni selección previa y han realizado la encuesta de forma totalmente voluntaria y anónima.

Además, por razones de limitaciones temporales, los cuestionarios han estado disponibles durante unos 5 meses, aproximadamente. Hemos detectado que cada vez que realizábamos un envío de invitaciones había un incremento puntual durante un periodo aproximado de una semana, pero que después la tasa de cuestionarios completados disminuía hasta el próximo envío de invitaciones.

Por tal razón, no creemos conveniente generalizar los resultados todavía. Nuestra intención es utilizar una muestra mucho más extensa y más representativa de todo tipo de usuarios finales, de diferentes perfiles y ámbitos, sea cual sea su formación con respecto a las nuevas tecnologías, para aumentar la capacidad de generalización de los resultados del estudio. También resultaría interesante extender este estudio a otros países (por ejemplo, a los de la Comunidad Europea, donde se está implantando el estándar HbbTV, o a otros países de habla hispana).

AGRADECIMIENTOS

El presente trabajo ha sido financiado, en parte, por el FEDER y por el MINECO, bajo el programa de apoyo a la I+D+i en el proyecto con referencia TEC2013-45492-R. También queremos agradecer la ayuda de Anne Moeskops, Isaac Almendros, Rebeca Díez, Fco. Javier Pastor, Dani Palacio, Francisco de Zulueta y Vicente Domingo Struch, por su inestimable ayuda en la elaboración de los cuestionarios on-line y el análisis de sus respuestas, así como a todos aquellos que han ayudado a difundir las invitaciones para participar en el estudio.

REFERENCIAS

- [1] Hybrid broadcast broadband TV, <http://www.hbbtv.org/>
- [2] Smart TV Alliance, <http://smarttv-alliance.org/>
- [3] M. Wijnants, W. Lamotte, J. De Meulenaere, W. Van den Broeck, "Qualitative Assessment of Contemporary Media Sharing Practices and Their Relationship to the sMS Platform", Hasselt University, IBBT-SMIT, 2012, pp. 31-36.
- [4] M. Montagud, F. Boronat, H. Stokking, R. van Brandenburg, "Inter-destination multimedia synchronization: schemes, use cases and standardization", Springer-Verlag, Multimedia Systems, 10 July 2012, pp. 459-482.
- [5] R. van Brandenburg, H. Stokking, O. van Deventer, F. Boronat, M. Montagut, K. Gross, "Inter-Destination Media Synchronization (IDMS) Using the RTP Control Protocol (RTCP)", RFC 7272, IETF, June 2014.
- [6] R. Stanton, Second Screen Revolutionizing the Television Experience, TATA Consultancy Services, White Paper, 2012, último acceso: 8 de mayo de 2015, en: http://www.tcs.com/SiteCollectionDocuments/White%20Papers/Consulting_Whitepaper_Second-Screen-Revolutionizing-Television-Experience-Part-1_0512-2.pdf
- [7] YUME, La etnia emergente: interacción de los consumidores hispanos en todas las pantallas, White Paper, 2014, disponible en: <http://www.yume.com/es/insights/video-publicitario-de-investigación>
- [8] Advanced Television, Survey: Smart TV UIs a problem, último acceso: 8 de mayo de 2015, en: <http://advanced-television.com/2015/01/15/industry-pros-favour-2nd-screen-over-smart-tv/>
- [9] C. Concolato, S. Thomas, R. Bouqueau, J. Le Feuvre, "Synchronized Delivery of Multimedia Content over Uncoordinated Broadcast Broadband Networks", ACM MMSys'12, February 22-24, 2012, Chapel Hill, North Carolina, USA, ISBN 9781450311311
- [10] L. Beloqui, P. O Flaithearta, H. Melvin, "Interactive Multi-source Media Synchronization for HbbTV", Media Synchronization Workshop 2012, October 11, 2012, Berlin, Germany, ISBN 9789059864108
- [11] F. Boronat, M. Montagut, H. Stokking, O. Niamut, "The need for Inter-Destination Synchronization for Emerging Social Interactive Multimedia Applications", IEEE Communications Magazine, November 2012, pp. 150-158.
- [12] The Office of Research Integrity (ORI), Descriptive study design, último acceso: 8 de mayo de 2015, en: http://ori.hhs.gov/education/products/sdsu/res_des1.htm

Herramientas TIC para la mejora del rendimiento y la planificación de redes PLC para *Smart Grids*

Miguel Seijo¹, Gregorio López¹, José Ignacio Moreno¹, Javier Matanza², Fernando Martín³, Cristina Martínez⁴

¹ Universidad Carlos III de Madrid, Dept. Ing. Telemática, nombre.apellido@uc3m.es

² ICAI, Grupo de Electrónica Automática y Comunicaciones, jmatanza@upcomillas.es

³ Unión Fenosa Distribución, fmartins@gasnatural.com

⁴ ZIV Metering Solutions, cristina.martinez@cgglobal.com

Resumen—Las redes de comunicaciones *Machine-to-Machine* (M2M) para *Smart Grids* deben ajustarse a requisitos muy exigentes tanto desde el punto de vista técnico (p. ej., baja latencia, alta disponibilidad) como económico (p. ej., bajos costes de despliegue y operación). Las tecnologías *Power Line Communications* (PLC) son especialmente relevantes en este tipo de entornos precisamente porque representan una buena solución de compromiso entre ambas perspectivas. Sin embargo, el cable eléctrico no deja de ser un medio de comunicaciones hostil, en el que se ha investigado mucho recientemente, pero que aún presenta retos tecnológicos.

Este artículo presenta las herramientas desarrolladas en el ámbito de los proyectos de investigación nacionales PRICE-GEN y OSIRIS para proporcionar soluciones a algunos de los retos tecnológicos más relevantes en este ámbito. Dichas herramientas se centran, concretamente, en la monitorización y diagnosis de redes *PowerLine Intelligent Metering Evolution* (PRIME) y en la simulación de redes PRIME y celdas PLC de banda ancha en redes eléctricas de media tensión (MV-BPL).

Palabras Clave—*Network Forensics, Machine-to-Machine Communications, Medium Voltage Broadband over Power Line, PowerLine Intelligent Metering Evolution, Simulación, Smart Grids*

I. INTRODUCCIÓN

Las redes eléctricas inteligentes (del inglés, *Smart Grids*) suponen una revolución especialmente para la distribución y el consumo de energía, ya que, por un lado, es en las redes de media y baja tensión donde se introducen los cambios más radicales (p. ej., penetración de generación distribuida basada en renovables, penetración de coche eléctrico, gestión activa de la demanda) y, por otro lado, la gestión de la red eléctrica tradicionalmente se ha llevado a cabo a gran escala, no llegando hasta el nivel de las redes de media y baja tensión.

Las redes de comunicaciones *Machine-to-Machine* (M2M) son de especial importancia para las *Smart Grids* en tanto en cuanto son clave para permitir la monitorización y control en casi tiempo real del elevado número de puntos de consumo y de generación previstos. Por lo tanto, deben cumplir requisitos muy exigentes desde el punto técnico (p.ej., baja latencia, alta disponibilidad), pero también desde el económico (bajos costes de despliegue y operación) [1]. Precisamente, las comunicaciones *PowerLine Communications* (PLC) son especialmente relevantes en este tipo de entornos porque se ajustan a dichos requisitos desde ambas perspectivas.

Así, las tecnologías PLC de banda estrecha (del inglés, *Narrowband PLC* (NB-PLC)) están siendo ampliamente desplegadas en infraestructuras de medición avanzada (AMI) para conectar los contadores inteligentes con los concentradores de datos situados en los centros de transformación

(CTs) [2]. Entre las distintas opciones disponibles en el mercado, destaca la tecnología *PowerLine Intelligent Metering Evolution* (PRIME), cuya especificación ha sido liderada por distribuidoras eléctricas nacionales, como Unión Fenosa o Iberdrola, y recientemente se ha convertido en un estándar internacional (ITU-T G.9904) [3].

Para comunicar los CTs con los sistemas de información, las comunicaciones PLC de banda ancha sobre la infraestructura eléctrica de media tensión (del inglés, *Medium Voltage Broadband over Power Line* (MV-BPL)) están ganando importancia como alternativa a otras soluciones, como por ejemplo las basadas en comunicaciones celulares [4]. En este tipo de redes, para el caso concreto de AMI, los concentradores de datos ubicados en los CTs se organizan formando lo que se conoce como celdas, para enviar la información a través de ellos mismos hasta un nodo que se conoce como Pasarela. Los dos estándares principales que especifican este tipo de comunicaciones son el IEEE 1901 y el ITU-T G.hn [5]. Aunque dichas especificaciones son distintas, en teoría, en una misma celda pueden co-existir equipos que utilizan una y otra.

Si bien es cierto que las comunicaciones basadas en la propia infraestructura eléctrica presentan multitud de ventajas para las distribuidoras eléctricas, también es cierto que se trata de un medio de comunicaciones hostil, en el que se ha investigado mucho recientemente, pero que aún presenta retos tecnológicos. Algunos de esos retos van en la línea de:

- Desarrollar **herramientas de monitorización** que permitan depurar y diagnosticar problemas en redes de comunicaciones PRIME en operación.
- Desarrollar **herramientas de simulación** que faciliten el despliegue y evaluación de redes basadas en PLC, minimizando riesgos y costes.

Este artículo se centra en las herramientas desarrolladas en el ámbito de los proyectos PRICE-GEN [6] y OSIRIS [7] para proporcionar soluciones a dichos retos tecnológicos.

El resto del artículo está organizado de la siguiente manera: en la sección II se describen las herramientas desarrolladas. En la sección III se presentan las infraestructuras de las que se dispone en el ámbito del proyecto OSIRIS para seguir mejorando y validando dichas herramientas. En la sección IV se resumen los principales resultados que se han obtenido con ellas hasta la fecha. La sección V presenta los trabajos futuros que se plantea acometer para seguir mejorando las herramientas haciendo uso de la infraestructura de la que

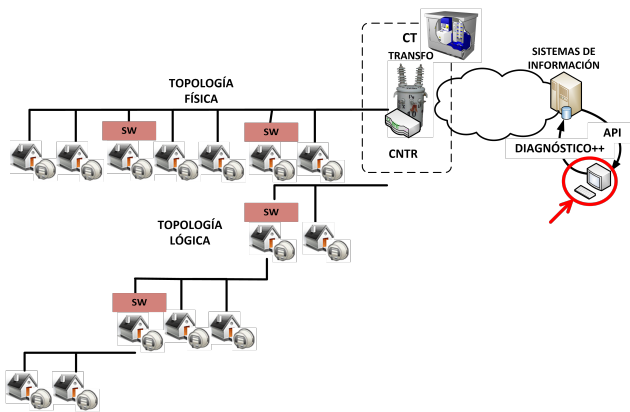


Fig. 1. Herramienta de diagnóstico de alto nivel empleada en los sistemas de información.

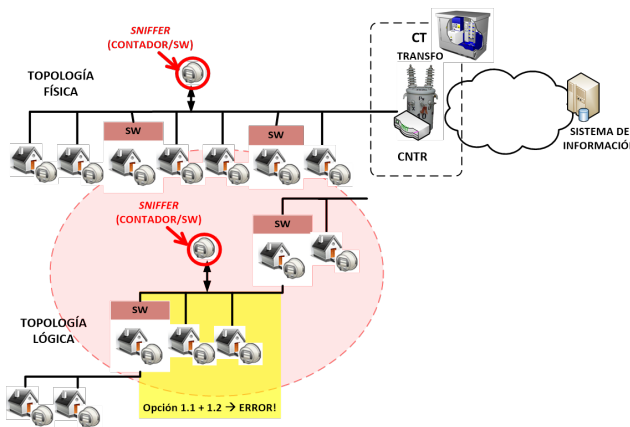


Fig. 2. Herramienta portátil para captura e inyección de tráfico.

se dispone. Finalmente, la sección VI concluye el artículo destacando sus principales contribuciones.

II. HERRAMIENTAS DESARROLLADAS

A. Herramientas para monitorización y diagnosis de redes PRIME

Las herramientas para monitorización y diagnosis de redes PRIME, se centran en dos focos de actuación:

- Una herramienta de diagnosis de alto nivel basada en el procesamiento de ficheros proporcionados por concentradores comerciales (p.ej., trazas de tráfico, ficheros de topología), accesibles a través de los distintos mecanismos que estos incluyen (interfaz web, telnet, etc.). El esquema mostrado en la Figura 1 ilustra el escenario de uso de esta herramienta.
- Una herramienta de captura e inyección de tráfico que pueda ser insertada en cualquier punto de la red eléctrica. Esta herramienta pretende aportar una solución móvil para la captura de trazas de tráfico y la diagnosis de problemas en base al análisis de las mismas para redes en producción. El esquema mostrado en la Figura 2 ilustra la zona de actuación de esta herramienta.

La primera opción permite acotar la rama donde se produce el error, mientras que la segunda opción se utilizaría en la rama afectada para obtener un diagnóstico detallado.

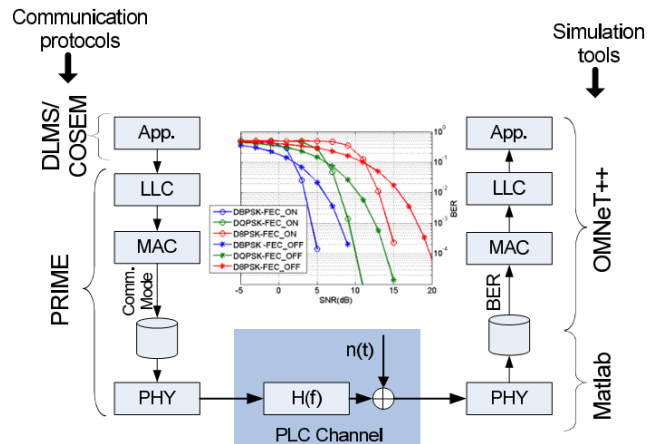


Fig. 3. Esquema de la arquitectura de simulación utilizada.

Para la herramienta de diagnóstico de alto nivel se ha desarrollado un analizador de traza hexadecimal que extrae determinados parámetros de comunicaciones, la relación señal a ruido (SNR) y los tiempos de lectura de trazas reales. Respecto a dicha herramienta, existe una versión implementada en MATLAB que ejecuta localmente los algoritmos y extrae los resultados y las representaciones significativas de los mismos [8]. Está en desarrollo una mejora de esta herramienta que se ejecute desde la web y evite la dependencia de herramientas comerciales como MATLAB, permitiendo importar trazas, procesarlas y visualizar los resultados.

La herramienta portátil de captura e inyección se encuentra actualmente en desarrollo y permitirá tanto la captura y almacenamiento en local como la captura en tiempo real de las trazas y su posterior procesamiento para diagnosticar problemas en secciones concretas de redes PRIME en operación.

B. Simulador de redes PRIME

La herramienta de simulación para evaluación y planificación de redes PRIME toma como punto de partida el simulador desarrollado en [9]. Este simulador combina MATLAB y OMNeT++ para modelar los efectos de la capa física (PHY) y de las capas superiores (*Media Access Control* (MAC), *Logical Link Control* (LLC) y aplicación) respectivamente. La Figura 3 ilustra el funcionamiento del simulador.

Respecto a la torre de protocolos, la implementación de las capas PHY, MAC y LLC se basa en el estándar PRIME; mientras que la capa de aplicación se modela en base a los *payloads* de los protocolos de nivel superior, como, por ejemplo, *Device Language Message Specification/COmpanion Specification for Energy Metering* (DLMS/COSEM) en el caso de AMI.

La interacción entre MATLAB y OMNeT++ se lleva a cabo de la siguiente manera: conocidos la potencia de transmisión y el nivel de ruido base cuando un nodo envía un mensaje a otro, se calcula la potencia recibida en base a una matriz de atenuaciones que contiene la atenuación entre cada par de nodos. Con la potencia recibida y la potencia de ruido base se obtiene la SNR, tal y como se muestra en la Ecuación 1. Conocidos la SNR y la constelación utilizada, se obtiene la tasa de error de bit (BER) mediante las curvas SNR frente a BER que se muestran en la Figura 3. Dichas curvas se calculan

previamente en base al estándar usando para ello MATLAB. OMNeT++ utiliza los valores de BER obtenidos para decidir si el mensaje recibido contiene errores (y por tanto debe ser descartado) o por el contrario fue recibido correctamente (y puede ser procesado por las capas superiores).

$$SNR[dB] = Pot_{Rx}[dBW] - Pot_{noise}[dBW] \quad (1)$$

$$= (Pot_{Tx}[dBW] - Att[dB]) - Pot_{noise}[dBW]$$

La versión original del simulador ha sido extendida permitiendo, por ejemplo, simular no sólo topologías aleatorias sino también topologías fijadas *a priori*. Esta nueva característica permite simular topologías de red reales obtenidas a partir del procesamiento del fichero de topología que proporcionan los concentradores (S11.xml). Además, no sólo es posible calcular las atenuaciones entre cada par de nodos en base a teoría de líneas de transmisión [10], sino que también pueden fijarse para forzar determinadas tasas de error para cada par de nodos de la red dependiendo del nivel lógico en el que se encuentren.

C. Simulador de celdas MV-BPL

La herramienta de simulación para evaluación y planificación de celdas MV-BPL se basa en el estándar IEEE1901 [11]. La versión actual de esta herramienta se decidió implementarla en MATLAB por los siguientes motivos:

- IEEE1901 utiliza *Time Division Multiple Access* (TDMA) para el intercambio de datos, lo que hace que las comunicaciones sean deterministas, reduciendo por tanto los beneficios que aportaría utilizar un simulador de redes de comunicaciones como OMNeT++.
- MATLAB está optimizado para trabajar con matrices, por lo que el manejo de las mismas es muy sencillo.
- MATLAB permite generar un fichero ejecutable que ofrece una interfaz gráfica de usuario de uso (ver Figura 4) sencillo que enmascara la complejidad de la herramienta.

La versión actual de esta herramienta calcula la latencia de ida y vuelta (RTT) entre todos los nodos de una celda MV-BPL dada y la Pasarela de la misma (parte izquierda de la Figura 4) en base a una serie de parámetros de entrada (parte derecha de la Figura 4).

Los parámetros de entrada que se consideran en la versión actual de la herramienta se describen a continuación:

- *Cell*: Este parámetro representa la celda MV-BPL a analizar, expresada en forma de fichero con las siguientes variables:
 - *ID*: Vector con los identificadores de todos los nodos que componen la celda MV-BPL.
 - *D*: Matriz de dimensiones NxN (donde N es el número de nodos de la celda MV-BPL) en la cual las filas representan la distancia (en metros y salto a salto) entre el nodo asociado a esa fila y la Pasarela.
 - *gw_with_two_bpl_interfaces*: Parámetro de tipo booleano para indicar si la Pasarela utiliza dos interfaces de comunicación MV-BPL (valor '1') o una única interfaz (valor '0')

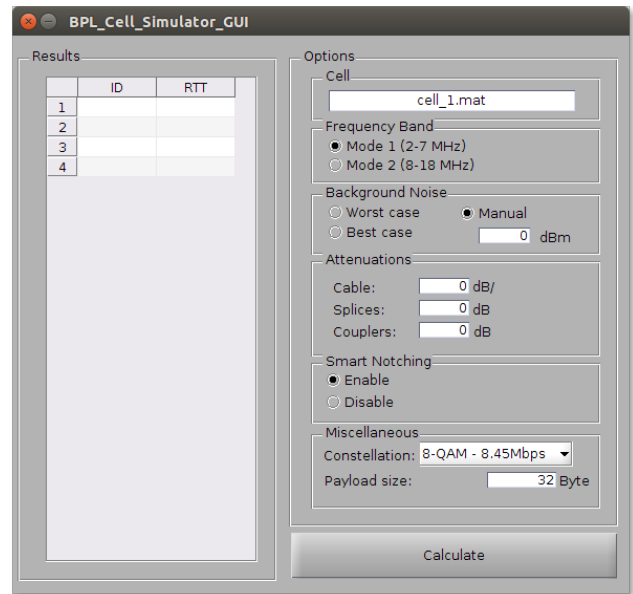


Fig. 4. Interfaz gráfica de usuario de la herramienta para evaluación y planificación de celdas MV-BPL

- *Frequency Band*: Es un parámetro que permite seleccionar la banda de frecuencia en la que opera la celda MV-BPL, distinguiendo entre el Modo 1 (entre 2 y 7 MHz) y Modo 2 (entre 8 y 18 MHz).
- *Background Noise*: Calcula la potencia de ruido de fondo en la banda seleccionada para el peor o mejor caso tal y como se especifica en el Anexo F “Channel and noise analysis and models for BPL systems” del estándar IEEE1901 [11]. La potencia de ruido también puede introducirse manualmente a través de la interfaz gráfica (en dBm).
- *Attenuations*: Se consideran tres fuentes distintas de atenuación:
 - *Cable*: Representa la atenuación del cable (en dB/m). Si su valor es '0', la atenuación se calcula usando la Ecuación 2 [12], donde f es la frecuencia en MHz y d es la distancia en metros. En la versión actual de la herramienta, f toma el valor central de la banda de frecuencia.

$$Cables = \mu_{MV}(f, d) = 1.77f + 0.01d + 32.9 \quad (2)$$
 - *Splices*: Representa la atenuación debida a los empalmes, que puede repercutir en pérdidas de transmisión bajas o considerables (> 10dB) [13]. En la versión actual de la herramienta, en el caso de que se especifique un valor para este parámetro, se asume un empalme por nodo.
 - *Couplers*: Representa la atenuación introducida por el método utilizado para acoplar el equipo de comunicaciones MV-BPL con el cable de media tensión. Distintos métodos de acoplo implican atenuaciones diferentes [14]. Además, dependiendo de si se utilizan acoplos inductivos o capacitivos, cambios en el estado de conmutación de la red de media tensión (MT) pueden conllevar atenuaciones de la señal de

hasta ± 20 dB [13].

- **Smart Notching:** Si se utiliza esta técnica para mitigar los efectos de las interferencias, se asume una penalización del 23% con respecto a la tasa de bits original según la Ecuación 3 [15].

$$tasa_bits_{ef} = tasa_bits * (1 - 0.23 * smart_notching) \quad (3)$$

- **Constellation:** permite elegir la modulación y su tasa de bits asociada (en bps) dependiendo de la banda de frecuencias seleccionada y en base a la sección 13.4 “PPDU structure and generation” del estándar IEEE1901 [11].
- **Payload Size:** permite especificar el tamaño de los datos a nivel de aplicación. Para calcular el tamaño de paquete a nivel PHY, se añaden 64 bytes de cabeceras al Payload size, en base a la Tabla 6-2 “General FFT MAC Frame Format” del estándar IEEE1901 [11]. El tamaño a nivel PHY es necesario para calcular el tiempo de transmisión, ya que las tasas de bit están calculadas también a nivel PHY.

Basándose en estos parámetros de entrada, la latencia de ida y vuelta entre cada nodo de la celda MV-BPL y la Pasarela se calcula mediante el procedimiento que se explica a continuación:

- 1) Se calculan las matrices de potencia transmitida para la comunicación de los nodos a la Pasarela, P_{tx_up} , y de la Pasarela a los nodos, P_{tx_down} . Para ello, primero se calcula la potencia de transmisión máxima p_{max} integrando la densidad espectral de potencia indicada en la hoja de características del equipamiento (en la versión actual de la herramienta, -50 dBm/Hz tal y como se especifica en la hoja de características del API-2000-SA de Ormazabal Current [16]) en todas las frecuencias que lleven tonos OFDM (ver Figura 13.20 “FFT transmit spectrum mask” del estándar IEEE1901 [11]) dentro de la banda de frecuencia seleccionada. Como ya se ha explicado en esta sección, la potencia de transmisión asociada a los repetidores de la celda será $p_{RP} = p_{max}/2$. Las matrices P_{tx_up} y P_{tx_down} se obtienen a partir de la matriz D.
- 2) Se calculan las matrices de potencia recibida para comunicaciones ascendentes y descendentes (P_{rx_up} y P_{rx_down} respectivamente). Estas matrices se calculan utilizando la Ecuación 4, donde Mask es una matriz NxN con ‘1’ en las posiciones donde la matriz D es distinta de ‘0’ y ‘0’ en el resto.

$$P_{rx_up/down}[dBm] = P_{tx_up/down}[dBm] - Cables \left[\frac{dB}{m} \right] * D[m] - Slices[dB] * Mask - Couplers[dB] \quad (4)$$

- 3) Se obtiene la matriz de potencia de ruido P_N multiplicando la potencia de ruido p_n por la matriz Mask anteriormente mencionada. Como ya se ha comentado, el parámetro p_n puede introducirse manualmente a través de la interfaz gráfica, o de lo contrario se calcula integrando la función de densidad espectral de potencia

especificada en el Anexo F “Channel and noise analysis and models for BPL systems” del estándar IEEE1901 [11] en todas las frecuencias cubiertas por tonos de la multiplexación *Orthogonal Frequency Division Multiplexing* (OFDM), lo que a su vez depende de la constelación y de la banda de frecuencia seleccionadas.

- 4) En base a las matrices de potencia recibida y de ruido, las matrices de SNR para las comunicaciones ascendentes y descendentes (SNR_{up} y SNR_{down} respectivamente) se calculan usando 5.

$$SNR_{up/down}[dB] = P_{rx_up/down}[dB] - P_N[dB] \quad (5)$$

- 5) Combinando $SNR_{up/down}$ y la Constelación seleccionada, las matrices de BER ($BER_{up/down}$) se obtienen en base a los procedimientos descritos en la sección 13.6 “Frame control forward error correction” del estándar IEEE1901 [11]. Conviene matizar que estas matrices incluyen la BER resultante tras aplicar el *Forward Error Correction* (FEC).
- 6) Basadas en las matrices $BER_{up/down}$, las matrices con la tasa de error de paquete (PER) se calculan utilizando la Ecuación 6

$$PER_{up/down} = 1 - (1 - BER_{up/down})^{packet_size} \quad (6)$$

- 7) Las matrices $Retx_{up}$ y $Retx_{down}$ se calculan basándose en $PER_{up/down}$. Estas matrices de dimensión NxN incluyen el número de retransmisiones por cada salto para cada comunicación ascendente o descendente. Notar que el número de retransmisiones cambia de una ejecución a otra, ya que dependen de la PER que caracteriza cada salto/enlace.
- 8) Las matrices de tiempo de transmisión para comunicaciones ascendentes o descendentes (T_{tx_up} y T_{tx_down} respectivamente) se calculan usando 7, donde: $t_{tx_up/down} = tamaño_paquete / tasa_bits_efectiva$.

$$T_{tx_up/down} = R_{tx_up/down} * t_{tx_up/down}[s] \quad (7)$$

- 9) La matriz de dimensiones NxN $T_{overall}$ incluye el tiempo total (considerando la comunicación ascendente y descendente) asociado a cada salto/enlace. Esta matriz se calcula usando la Ecuación 8, donde puede verse que se considera incluso el tiempo de propagación ($T_{prop}[s] = \frac{D[m]}{c[m/s]}$), aunque se comprueba que es despreciable comparado con el tiempo de transmisión.

$$T_{overall} = T_{tx_up}[s] + T_{tx_down}[s] + 2 * T_{prop}[s] \quad (8)$$

- 10) Finalmente, la latencia entre cada uno de los nodos de la celda MV-BPL y la Pasarela es un vector Nx1 que resulta de sumar los elementos de cada fila de $T_{overall}$.

III. INFRAESTRUCTURAS SOPORTE PARA DESARROLLO Y VALIDACIÓN

A. Panel de teledidada

El panel de teledidada instalado en la UC3M, se ha montado principalmente como soporte a la creación y validación de las herramientas PRIME previamente descritas.

El panel se compone de:

- Cuatro contadores monofásicos de la marca ZIV, modelo 5CTME2C47536AVF, los cuales disponen de interfaz de

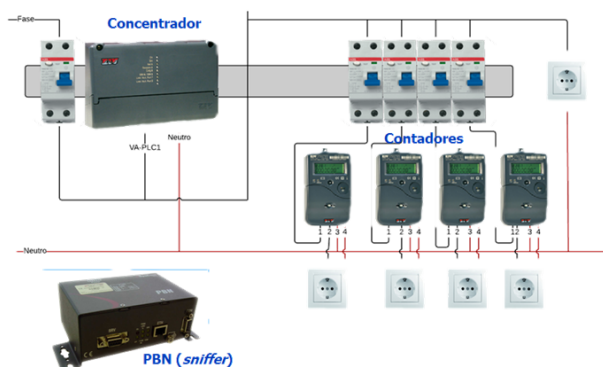


Fig. 5. Panel de contadores.

comunicaciones PLC bajo el estándar PRIME

- Un concentrador de comunicaciones PRIME PLC trifásico de la marca ZIV, modelo 4CCT-EA8-032126BJ
- Cuatro interruptores de control de potencia (ICPs) bipolares de 16A
- Un ICP bipolar de 20A
- Cinco bases de enchufe

Todo el panel va alimentado mediante un enchufe monofásico y se pueden conectar cargas directamente a la línea principal del concentrador o a las líneas provenientes de cada uno de los contadores, ya sea para conectar nuevos equipos de comunicaciones PRIME o bien conectar cargas en cada una de las líneas de cada contador de modo que se puedan medir consumos. Los distintos ICP permiten conectar y desconectar el panel por completo o bien cada uno de los contadores y sus tomas de enchufe correspondientes.

La Figura 5 muestra el esquema de conexionado del panel. El concentrador va además conectado por medio de un cable Ethernet a una red IP para su gestión. Es importante recalcar que el panel está diseñado con fines de investigación en el área de las comunicaciones, y por tanto dispone del interconexionado mínimo necesario para la comunicación entre los contadores y el concentrador.

La gestión y el envío de solicitudes desde el concentrador se realiza en remoto por medio de interfaz web, mientras que las tareas avanzadas como la captura de tráfico o el análisis de espectro se realizan utilizando el software *ZIV PRIME Manager*, el cual se conecta al concentrador remotamente a través de la interfaz Ethernet del mismo.

Además de los elementos anteriormente descritos, y en el contexto del proyecto OSIRIS, el fabricante de equipos de comunicaciones ZIV ha facilitado el equipo *PRIME Base Node* (PBN) que puede actuar como nodo base, nodo de servicio o monitor de redes PRIME (ver Figura 5) y puede conectarse a cualquiera de las tomas de enchufe disponibles en el panel.

B. LINTER

La colaboración con Unión Fenosa Distribución (UFD) como parte del consorcio del proyecto Optimización de la Supervisión Inteligente de la Red de dIstribución (OSIRIS) incluye la realización de pruebas en el Laboratorio de Integración de Redes (LINTER), su laboratorio de redes eléctricas inteligentes.

Los objetivos principales del LINTER de UFD son los siguientes:

- Proporcionar un entorno interno de formación para profesionales de ámbitos relacionados con las redes eléctricas inteligentes.
- Realizar tests sobre el nuevo equipamiento de manera previa a su instalación en campo para asegurar su integración y funcionalidad.
- Realizar pruebas sobre las nuevas versiones de *hardware* y *software* antes de actualizar los equipos en servicio.
- Analizar incidentes mediante la creación de escenarios en el entorno de laboratorio que permitan investigar las causas de las anomalías detectadas en campo.
- Evaluar el rendimiento del equipamiento en los distintos escenarios posibles de cargas eléctricas y variables de comunicaciones.
- Probar situaciones reales de campo, usando atenuadores e interruptores para simular cargas y distancias, cargas variables según el ruido e incluso generadores de ruido.

Para conseguir estos objetivos, el laboratorio está equipado con contadores inteligentes y concentradores de los principales fabricantes del mercado, así como con instalaciones de micro-generación y puntos de carga de vehículo eléctrico. [17].

C. Celdas MV-BPL

La colaboración con UFD como parte del consorcio del proyecto OSIRIS también permite la obtención de parámetros de red de distintas celdas MV-BPL desplegadas.

En concreto, los escenarios puestos a nuestra disposición comprenden una celda formada por 7 CTs unidos por cable eléctrico de longitud inferior a 1 Km y otra celda formada por 2 CTs unidos por cable eléctrico tanto aéreo como subterráneo, con una longitud total inferior a 1 Km.

IV. RESULTADOS OBTENIDOS

A. Herramientas para monitorización y diagnosis de redes PRIME

De la herramienta de monitorización de redes PRIME, se obtuvieron resultados a partir de las trazas de mensajes intercambiados entre los contadores y el concentrador. Estas trazas, facilitadas por el LINTER, fueron extraídas de un concentrador de Circutor e incluyen valores separados por comas para los parámetros de la capa PRIME y representación hexadecimal del protocolo DLMS/COSEM.

Tras analizar la traza con la herramienta desarrollada, se obtuvieron indicadores de calidad de las comunicaciones (en concreto, el tiempo de lectura (TTR) de cada contador), parámetros de implementación de PRIME y DLMS/COSEM (la *Maximum Transfer Unit* (MTU) y el tamaño de ventana efectiva), además de otra información como los valores de SNR de los contadores con el concentrador y la modulación utilizada.

Del análisis de la modulación utilizada, se observó que los fabricantes fuerzan el uso del modo de comunicación más robusto, (*Differential Binary Phase Shifting Keying* (DBPSK) con FEC activado). Este modo permite tasas de transferencia de hasta 21,4 Kbps.

Los resultados del análisis del TTR se muestran en la Figura 6. Se puede observar que la mayoría de los contadores

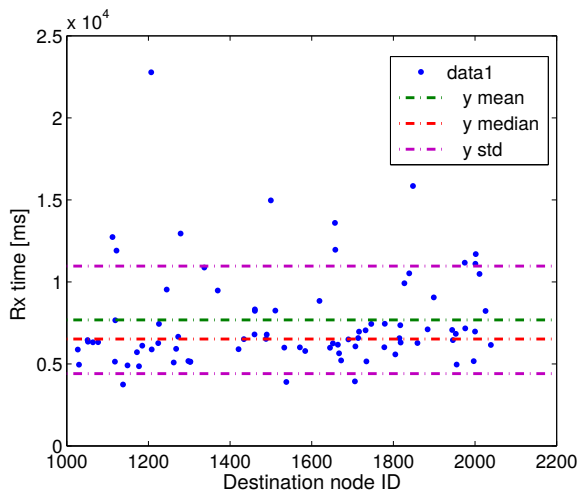


Fig. 6. TTR_i - ID de nodo vs Tiempo

necesitan una cantidad similar de tiempo para ser leídos, ya que sus valores de TTR se encuentran cercanos a la mediana. Sin embargo, también se observan algunos valores atípicos.

Del estudio de los parámetros de implementación de los protocolos PRIME y DLMS/COSEM (MTU y tamaño de ventana efectiva), se obtienen 13 grupos con distintas configuraciones, mostrados en la Tabla I. Se observa que estos grupos no se ajustan necesariamente a los distintos fabricantes de contadores existentes en el escenario analizado (9). Esto puede deberse a la coexistencia de distintas versiones de *firmware* en distintos contadores de un mismo fabricante, tanto del *firmware* de la capa PRIME como de la capa DLMS/COSEM, ya que la MTU depende exclusivamente de la capa PRIME, pero el tamaño de ventana efectiva calculado por esta herramienta tiene en cuenta tanto parámetros de PRIME (bit de *FLUSH*) como de DLMS/COSEM (valor de *NSEGS*). La falta de correlación entre los grupos y los fabricantes de equipos también puede deberse a que los equipos modifiquen estos parámetros según las condiciones de la red, aunque para este caso las variaciones deberían ser bajas, debido a la lectura secuencial de los contadores, que equipara la congestión de la red para los distintos equipos, y a las condiciones ideales de laboratorio en las que fue extraída la traza, que equiparan las condiciones físicas de conexión entre los distintos contadores utilizados.

B. Simulador de redes PRIME

Utilizando la herramienta desarrollada se han realizado distintas simulaciones tomando como referencia una topología de red real obtenida de una prueba realizada en el LINTER. Dicha topología comprende 91 contadores inteligentes, 3 de los cuales funcionan como repetidores, 2 de ellos con un único “hijo” y 1 de ellos con 2 “hijos”.

Para cada una de estas simulaciones, se han modificado dos tipos de parámetros: (1) parámetros de nivel de enlace como la MTU y el tamaño de ventana; y (2) parámetros físicos del canal, como la atenuación del cable para obtener unas BER determinadas. Asimismo, DBPSK con FEC ON se fijó como modulación, al comprobarse que es la implementada por los contadores actualmente.

Tabla I

TABLA DE CONFIGURACIONES DE MTU Y VENTANA EFECTIVA

# contadores	Ventana efectiva	MTU (Bytes)
7	6	47
2	1	211
33	4	67
5	4	64
4	2	67
2	1	67
18	1	115
4	4	68
3	3	65
1	1	178
2	3	52
1	1	79
1	1	178

Los resultados son evaluados en base al tiempo necesario para la lectura de todos los contadores (TTRAll), utilizando MTU de 50, 100, 200 y 256 bytes; *Window Size* (WS) de 1, 2, 5 y 10; y atenuaciones en el cable correspondientes al caso ideal sin errores de transmisión (Escenario 1), a un caso intermedio sin errores de transmisión entre contadores del mismo nivel y con errores entre dispositivos de niveles contiguos (Escenario 2), y a un caso peor con errores de transmisión en ambos casos (Escenario 3).

Las simulaciones se han repetido para cada combinación de (MTU, WS, escenario) un número de veces lo suficientemente alto como para poder calcular intervalos de confianza al 95% utilizando para ello la Ecuación 9, donde μ es la media muestral, σ la desviación estándar de la muestra y n el tamaño de la muestra. La Tabla II resume los intervalos de confianza calculados.

$$CI = \mu \pm \left(1.96 \cdot \frac{\sigma}{\sqrt{n}} \right) \tag{9}$$

Analizando la dependencia del TTRAll con respecto a la MTU, puede observarse como un incremento de la MTU produce una disminución del TTRAll en el caso sin errores (Escenario 1). La misma tendencia para valores extremos se observa en el Escenario 2. Para el Escenario 3, sin embargo, la tendencia cambia. Esto se debe a que, en canales sin errores, aumentar la MTU reduce el *overhead*, mientras que, en canales con errores, incrementar la MTU, incrementa la PER y, además, en caso de producirse un error, el tiempo de retransmisión también es mayor. Por lo tanto, la MTU debe elegirse teniendo en cuenta las condiciones del canal.

En cuanto a la dependencia del TTRAll con respecto al WS, para todos los escenarios, un incremento en el WS tiene como consecuencia un decremento en el TTRAll. Esto sucede porque el simulador implementa el mecanismo de Rechazo Simple (como se indica en las especificaciones de PRIME), donde sólo los paquetes enviados tras el último paquete asentido son reenviados. Por lo tanto, incrementar el WS siempre se traduce en un mejor rendimiento de la red en términos de latencia.

C. Simulador de celdas MV-BPL

La validación preliminar de esta herramienta se llevó a cabo en el ámbito del proyecto PRICE-GEN. Para ello se utilizaron

Tabla II
INTERVALOS DE CONFIANZA AL 95% DE TTRALL (EN S) EN FUNCIÓN DEL WS Y LA MTU PARA LOS TRES ESCENARIOS CONSIDERADOS

MTU (Bytes)	WS	Escenario 1 (BER11=BER12=0)				Escenario 2 (BER21=0;BER22=1.67 · 10 ⁻⁴)				Escenario 3 (BER31=5 · 10 ⁻⁵ ; BER32=7 · 10 ⁻⁴)			
		50	100	200	256	50	100	200	256	50	100	200	256
1	1	424.12	242.32	160.51	141.51	972.77	786.30	669.97	556.81	1447.32	1425.18	1280.25	1509.85
		426.12	243.37	161.12	142.12	1109.79	868.93	732.76	642.44	1722.37	1615.26	1546.68	1808.63
	2	328.18	200.28	139.76	126.87	619.16	487.02	426.36	434.68	1284.41	1108.57	1121.77	1236.93
		329.80	201.10	142.13	128.29	650.98	520.25	461.49	471.19	1463.85	1348.60	1391.35	1487.03
5	5	279.02	172.56	126.19	114.56	432.36	328.76	297.61	275.05	814.04	833.85	990.81	1033.62
		280.16	175.28	127.81	116.59	467.80	351.61	321.64	302.56	902.29	985.82	1095.70	1251.90
10	10	259.08	162.74	119.27	114.56	342.55	278.23	225.93	275.15	681.97	808.29	795.96	1060.58
		259.94	166.85	119.71	116.59	362.80	297.35	242.35	302.63	745.60	898.89	877.97	1275.47

dos celdas MV-BPL reales operadas por Iberdrola. La métrica utilizada para validar la herramienta fue la latencia de ida y vuelta entre el nodo más alejado de la Pasarela de una celda MV-BPL y la propia Pasarela.

Para ello, en primer lugar se midió la latencia de ida y vuelta entre dichos nodos y las Pasarelas de las respectivas celdas MV-BPL ejecutando un ping de 32 bytes 200 veces. Se asumió por tanto que las latencias de ida y vuelta medidas podían ajustarse a una distribución Gaussiana con media μ igual a la media muestral y desviación típica σ , estimada como la desviación típica de las muestras, y se utilizó la Ecuación 9 para calcular intervalos de confianza al 95%.

Sin embargo, ciertos parámetros de entrada de la herramienta, como el ruido, la constelación o la atenuación de los cables de MT, no eran conocidos para las celdas MV-BPL sujetas a estudio. La herramienta desarrollada considera multitud de parámetros de entrada que pueden tomar un gran número de valores distintos, lo que significa un número aún mayor de combinaciones. Por lo tanto, intentar alcanzar un valor de latencia cercano a los medidos mediante búsqueda por fuerza bruta no era una opción computacionalmente asequible, además de correr el riesgo de que si la configuración inicial no era la apropiada, el resultado podría no converger a un valor óptimo o incluso podría divergir. Por esta razón, la mayoría de los parámetros de entrada se fijaron de forma heurística y finalmente se realizó un ajuste fino variando la potencia de ruido entre el mejor y el peor caso especificados en el estándar, obteniéndose unos valores de latencia relativamente cercanos a los esperados. Para las latencias simuladas también se calcularon intervalos de confianza al 95% utilizando la Ecuación 9 y $n = 10.000$.

La Tabla III resume los parámetros de entrada y salida de la herramienta junto a los intervalos de confianza obtenidos para los tests realizados en las celdas MV-BPL reales. Como se puede observar, los intervalos de confianza simulados y experimentales no se solapan, lo que indica que los resultados difícilmente podrían coincidir. Sin embargo, el que los valores de entrada del simulador no sean conocidos resta valor a esta conclusión estadística.

V. TRABAJOS FUTUROS

A. Herramientas para monitorización y diagnosis de redes PRIME

En la tarea de monitorización el trabajo a realizar se apoya en el panel de teledirigida y en los recursos que nos ofrece el LINTER, tanto para la herramienta de diagnosis de alto nivel como para la herramienta de captura e inyección de tráfico.

Tabla III
RESUMEN DE LOS PARÁMETROS DE ENTRADA Y SALIDA UTILIZADOS JUNTO A LOS INTERVALOS DE CONFIANZA CALCULADOS.

Parámetros	Celda 1	Celda 2	
Entrada	Banda	Modo 1	
	p_n	-39.65 dBm	-48.2 dBm
	Att. Cable	Eq 2	Eq 2
	Att. Empalmes	0 dB	0 dB
	Att. Acoplos	0 dB	0 dB
	Smart Notching	No	Sí
	Constelación	8-QAM	16-QAM
Tamaño de paquete	32 Bytes	32 Bytes	
Salida	RTT	(35.1 , 35.8) ms	(145.2 , 148.3) ms
	RTT de celda real	(29.65 , 31.15) ms	(119.75 , 136.22) ms

Para el desarrollo de la herramienta de diagnóstico de alto nivel es necesario conocer en profundidad el conjunto de utilidades que proporciona el concentrador que aporte datos sobre el estado de la red, así como las interfaces que permitan la obtención de estos datos para crear una herramienta que sea capaz de utilizarlos para diagnosticar y localizar los problemas de la red. Para llevar a cabo este análisis en profundidad de las utilidades del concentrador es necesario disponer de un entorno de laboratorio o panel de pruebas, como el descrito en la sección III.A, donde tener contacto con el dispositivo y llevar a cabo la investigación y los desarrollos preliminares.

Para el desarrollo de la herramienta de captura e inyección de tráfico, se precisa de una red de pruebas a la que conectarse físicamente para realizar los test de los distintos modos de funcionamiento, utilizando la herramienta como nodo base, nodo de servicio y como monitor. También se precisa la conexión física a la red para desarrollar las tareas de decodificación de cabeceras, inyección de tráfico y análisis tanto del tráfico como de los parámetros físicos. El panel de pruebas se adecúa perfectamente a esta necesidad, siendo imprescindible para la implementación de las funcionalidades requeridas y los primeros tests en entorno controlado, mientras que el LINTER es de utilidad para la realización de pruebas en distintos escenarios y llevar a cabo una primera fase de validación antes de probar la herramienta en campo.

B. Simulador de redes PRIME

Para el simulador PRIME, el trabajo a realizar se centra en:

- Extraer los valores de SNR directamente a partir del análisis de trazas reales.
- Comparar los resultados de tiempo de lectura medido y simulado por nodo en lugar de para todos los contadores que forman una red PRIME.
- Implementar configuraciones más complejas del payload

que permitan simular otras órdenes y aplicaciones, más allá de la teledatada de consumos de los contadores.

- Simular CTs problemáticos (p. ej., con más de 1000 contadores).
- Analizar distintas alternativas que permitan modelar escenarios no existentes para su posterior simulación (p. ej., (1) realizando un barrido de combinaciones para una serie de parámetros de nivel físico y de red; (2) utilizando procesado de trazas para modelar escenarios tipo).
- Evaluar el posicionamiento óptimo de repetidores o la necesidad de utilizarlos en escenarios concretos.
- Mejorar la presentación y visualización de los resultados extraídos del simulador.

El panel de teledatada juega un papel crucial de cara a obtener trazas que nos permitan encontrar el método óptimo para extraer el SNR y para modelar los payload para otras órdenes y aplicaciones, cuyos resultados podrían ser además validados en el mismo panel para topologías sencillas. Mediante las distintas pruebas en el LINTER se extrapolarán las pruebas del panel a escenarios más complejos, permitiendo además realizar pruebas de laboratorio que ayuden a evaluar el posicionamiento de repetidores y la identificación de escenarios tipo.

C. Simulador de celdas MV-BPL

En base a los resultados presentados en la sección IV.C, una de las principales líneas de actuación en este caso consiste en validar la herramienta utilizando parámetros de entrada reales (p.ej., constelación utilizada, SNR en los distintos CTs, atenuaciones introducidas por empalmes y acopladores, etc.) medidos en las celdas MV-BPL presentadas en la sección III.C.

No obstante, también se contemplan líneas de trabajo futuras orientadas a la mejora de la herramienta en sí misma, como por ejemplo:

- Considerar distintas potencias de ruido en cada nodo de la celda.
- Calcular la atenuación de la línea de manera dependiente de la frecuencia [18].
- Implementar modulaciones adaptativas, en las que cada portadora OFDM utilice una constelación distinta dependiendo de las condiciones concretas de la sub-banda de frecuencia correspondiente.
- Considerar distintas tasas de transmisión para cada sentido de la comunicación de cada enlace (enlaces asimétricos).
- Validar la herramienta con parámetros de entrada medidos en celdas reales.
- Migrar la herramienta a una plataforma libre de licencias comerciales para eliminar la dependencia de Matlab (p. ej., Python) y hacerla accesible a través de la Web.

VI. CONCLUSIONES

Este artículo presenta un conjunto de herramientas TIC situadas en la vanguardia tecnológica en el área de las redes PLC, en tanto en cuanto proporcionan soluciones a algunos de los problemas más relevantes que se plantean al utilizarlas para la supervisión de redes de distribución eléctrica en entornos de *Smart Grid*. Concretamente, se presentan, por un lado, herramientas enfocadas a la diagnosis de problemas

en redes PRIME operativas en base a la monitorización de las mismas. Por otra parte, se describen herramientas de simulación para redes PRIME y celdas MV-BPL, cuyo principal objetivo es facilitar la planificación y evaluación de este tipo de redes. Además, el artículo también presenta las infraestructuras soporte para el desarrollo y validación de dichas herramientas, los resultados obtenidos con ellas hasta la fecha, así como las líneas de trabajo para su mejora.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad a través del Programa Estatal de I+D+i Orientado a los Retos de la Sociedad dentro del proyecto OSIRIS (RTC-2014-1556-3).

REFERENCIAS

- [1] G. López, J. I. Moreno, H. Amarís, F. Salazar, "Paving the road toward Smart Grids through large-scale advanced metering infrastructures", *Electric Power Systems Research*, vol. 120, pp. 194–205, 2015.
- [2] A. Haidine, A. Tabone, and J. Muller, "Deployment of power line communication by european utilities in advanced metering infrastructure," in the 17th IEEE International Symposium on Power Line Communications and its Applications (ISPLC), 2013, pp. 126–130.
- [3] ITU-T estándar G.9904. "Narrowband orthogonal frequency division multiplexing power line communication transceivers for PRIME networks". 2012
- [4] A. Sendin, J. Simon, I. Urrutia, I. Berganza, "PLC deployment and architecture for Smart Grid applications in Iberdrola," in the 18th IEEE International Symposium on Power Line Communications and its Applications (ISPLC), 2014, pp. 173–178.
- [5] S. Galli, A. Scaglione, Z. Wang, "For the Grid and Through the Grid: The Role of Power Line Communications in the Smart Grid", *Proceedings of the IEEE*, Vol. 99, No. 6, pp. 998–1027, June 2011.
- [6] Web oficial del proyecto PRICE: <http://www.priceproject.es/en>
- [7] Web oficial del proyecto OSIRIS: <http://www.unionfenosadistribucion.com/en/smart+grids/research+and+development/spanish/1297262412251/osiris.html>
- [8] M. Seijo, G. López, J. Matanza, J.I. Moreno, F. Martín, S. Alexandres, C. Rodríguez-Morcillo, "Let there be light: Unveiling how PRIME networks actually work" aceptado en ACM E-Energy 2015, Julio 2015.
- [9] J. Matanza, "Improvements in the PLC Systems for Smart Grids Environments", Tesis Doctoral.
- [10] O. G. Hooijen, "A channel model for the residential power circuit used as a digital communications medium" *IEEE Transactions on Electromagnetic Compatibility*, 1998, vol. 40, no. 4, pp. 331–336.
- [11] IEEE1901. "IEEE Standard for Broadband over Power Line Networks: Medium Access Control and Physical Layer Specifications". 2010.
- [12] S. Yoon, S. Jang, Y. Kim, S. Bahk, "Opportunistic Routing for Smart Grid With Power Line Communication Access Networks" *IEEE Transactions on Smart Grid*, Vol 5, No 1, 2014.
- [13] S. Dominiak, L. Andersson, M. Maurer, A. Sendin, I. Berganza, "Challenges of Broadband PLC for Medium Voltage Smart Grid Applications", 6th Workshop on Power Line Communications, 2012.
- [14] P. Mulroy, I. Gilbert, "Comparison of Coupling Methods in MV Equipment for Powerline Communications", 21st International Conference on Electricity Distribution (CIRED), 2011.
- [15] A. Schwager, "Powerline Communications: Significant Technologies to become Ready for Integration". Tesis Doctoral. 2010.
- [16] Current MV-BPL - API 2000 SA Product Datasheet. On-line: <http://www4.ormazabal.com/sites/default/files/ormazabal/productos/descargables/API-2000-SA.pdf>
- [17] Unión Fenosa Distribución, Web oficial del LINTER: <http://www.unionfenosadistribucion.com/en/smart+grids/1297137260045/find+out+about+our+laboratory.html>
- [18] A. G. Lazaropoulos, "Broadband over Power Lines Systems Convergence: Multiple-Input Multiple-Output Communications Analysis of Overhead and Underground Low-Voltage and Medium-Voltage BPL Networks", *ISRN Power Engineering*, Volume 2013 (2013), Article ID 517940, 30 pages.

V2I en campo cercano: modelado de las comunicaciones Eurobaliza-Tren para análisis de interferencias a sistemas de señalización ferroviaria

Lara Rodríguez¹, Christian Pinedo¹, Igor Lopez¹, Marina Aguado¹, Jasone Astorga¹, Marivi Higuero¹, Iñigo Adin², Guillermo Bisue²

¹Universidad del País Vasco UPV/EHU

²CEIT & TECNUM (Universidad de Navarra)

{lara.rodriguez, christian.pinedo, igor.lopez, marina.aguado, jasone.astorga, marivi.higuero}@ehu.eus,
{iadin, gbistue}@ceit.es

Resumen- El desarrollo del sector ferroviario depende en gran medida del despliegue de avanzados sistemas de señalización ferroviaria. Estos sistemas de señalización se soportan sobre arquitecturas de comunicación sujetas a un entorno electromagnético complejo. Este trabajo se enmarca en el contexto del desarrollo de herramientas que permitan acelerar el despliegue de estos nuevos sistemas de señalización facilitando el análisis del comportamiento de estos sistemas frente a interferencias. Concretamente, en este artículo presentamos el modelado en herramienta de simulación de propósito general de las comunicaciones eurobaliza-tren. Se trata de comunicaciones críticas puesto que pérdidas en la recepción pueden ocasionar paradas de emergencia del tren y fallos de disponibilidad. Para su correcto modelado se ha contado con resultados de pruebas de laboratorio con equipos reales. Se han obtenido indicadores de rendimiento de la capa física tales como la potencia recibida, SNR y tasa BER. Se trata de una etapa necesaria con el objetivo final de proporcionar indicadores de calidad de servicio ante diferentes escenarios de perturbación.

Palabras Clave- ERTMS, eurobaliza, BTM, simulador, interferencias

I. INTRODUCCIÓN

El impulso hacia una sociedad más sostenible recae en gran medida en el uso de los sistemas de transporte público guiado como el ferroviario. En estos entornos, los sistemas de señalización ferroviaria son un elemento clave. Estos sistemas de señalización hacen uso de tecnologías de comunicación ferroviarias y arquitecturas específicas sujetas a un entorno electromagnético muy complejo.

En el entorno europeo de alta velocidad, y en continua expansión hacia mallas regionales y también hacia el mercado no europeo, el sistema de señalización es el European Rail Traffic Management System (ERTMS).

Actualmente, el sistema ERTMS consta de dos subsistemas que permiten la interoperabilidad de los trenes: European Train Control System (ETCS), dedicado al control de seguridad de la circulación de los trenes, y GSM-R, dedicado a la comunicación por radio entre los trenes y los centros de control. En la actualidad numerosas operadoras ferroviarias se encuentran en fase de despliegue de ERTMS en alguno de sus niveles. Con el objetivo de acortar significativamente las actividades del ciclo de implementación (estudios, pruebas, validación), las herramientas de análisis y de simulación in-lab juegan un papel crítico.

El trabajo aquí presentado se encuadra en el contexto del

desarrollo de un entorno de simulación ERTMS que posibilite el análisis de las interferencias en sistemas de señalización. Concretamente, en este artículo presentamos el modelado de un elemento clave en el sistema ERTMS: la comunicación eurobaliza-tren.

El modelado se ha realizado sobre una plataforma de simulación de propósito general a la que se le han añadido las librerías y nuevos nodos necesarios para la correcta caracterización del sistema propuesto.

Para validar nuestro desarrollo se ha contado con medidas reales de la comunicación entre las balizas y el tren. En este artículo se proporciona el detalle del modelado de la comunicación eurobaliza-tren y también se comprueba que los resultados obtenidos en plataforma de simulación para parámetros tales como potencia, SNR y BER son coherentes con las medidas reales. Este desarrollo, al integrarse en un *framework* de simulación ferroviaria [1] que modela la pila completa de protocolos incluida la capa de aplicación ETCS, posibilita el estudio de indicadores de rendimiento de capas superiores ante interferencias. El modelado de dichas interferencias y su posterior análisis es objeto de trabajos futuros.

El documento está organizado de la siguiente manera: en la sección II se describen los elementos utilizados en el sistema ETCS para la comunicación vía-tren; en la sección III se presenta la metodología utilizada en el simulador para modelar la comunicación entre la vía y el tren; por último, en la sección IV se muestran las pruebas realizadas y los resultados obtenidos.

II. SISTEMA DE SEÑALIZACIÓN FERROVIARIO EUROPEO

El sistema ERTMS está compuesto de dos subsistemas [2]: el sistema a bordo (On-board subsystem) y el sistema en la vía (Trackside subsystem). Dentro de cada uno de estos sistemas existen diferentes elementos, cuya implantación en un entorno ferroviario depende del nivel de aplicación del sistema ERTMS.

Actualmente, ERTMS puede ser configurado para operar en los siguientes niveles:

- Nivel STM: El tren está equipado con el sistema ERTMS y opera en una línea con un sistema de seguridad nacional con interfaz STM. En este nivel de operación, las balizas de sistemas de seguridad nacionales, tales como ASFA o LZB, son leídas y su

telegramas con la información, formato, modulación y tasa de transmisión adecuadas.

Mientras dura el periodo de activación, la baliza emite el mismo telegrama constantemente. Así, en función de la velocidad del tren, la baliza se activa durante un mayor o menor periodo de tiempo, es decir, a mayor velocidad del tren el tiempo de activación de la baliza es menor. Durante este tiempo de activación se emitirán tantas veces el telegrama como sea posible.

Además, si se usan telegramas cortos (341 bits) se emitirá más veces que si se usan telegramas largos (1023 bits). El tren recibe el telegrama repetido un determinado número de veces en función de los factores anteriores y le basta con recibirlo una vez sin errores para poder procesar la información que contiene. De hecho, la emisión repetida del telegrama es una forma de fortalecer las comunicaciones baliza a BTM, ya que la comunicación baliza a BTM sólo falla si no se recibe ni una sola repetición del telegrama de forma correcta.

Una vez que el tren se encuentre a cierta distancia, la baliza se desactiva [6].

Para una correcta transmisión de la información, la baliza deberá cumplir los siguientes requerimientos definidos en la especificación de ERTMS [3] [4]:

- Banda de frecuencia de transmisión: de 3,951MHz a 4,516MHz.
- Bit rate de transmisión: 564,48 kbits/s
- Modulación: CPFSK
- Zona/Volumen de activación de la eurobaliza ante la presencia del tren:
 - $-1300 \text{ mm} < X < +1300 \text{ mm}$
 - $-1400 \text{ mm} < Y < +1400 \text{ mm}$
 - $+220 \text{ mm} < Z < +460 \text{ mm}$
- Longitud de los telegramas: 1023 o 341 bits

B. BTM

El tren recibe la información proporcionada por la eurobaliza a través del módulo BTM. Este módulo es el encargado de activar la baliza para que ésta comience a transmitirle telegramas. A su vez, se encargara de filtrar, demodular y decodificar la señal recibida y enviarla a otros módulos del sistema de a bordo para las funciones de control.

La comunicación tren-vía es una parte fundamental de los sistemas ferroviarios. Si un tren no recibe información de balizas en localizaciones en las que se esperan que estén, a falta de información en cuanto a velocidad, posicionamiento y/o curva de frenado, el tren ejecutará una parada de emergencia. Este mecanismo de seguridad permite salvar vidas, pero a su vez supone grandes pérdidas para los operadores ferroviarios, reduciendo la disponibilidad de la red y aumentando los costes operacionales. Por ello es necesario tener en cuenta los posibles factores externos que pueden afectar y/o inhabilitar dichas comunicaciones. Este documento recoge un análisis del comportamiento del canal de comunicación entre baliza y tren.

III. MODELADO EN EL SIMULADOR

Para el estudio del canal de comunicación entre la vía y el tren se ha empleado la herramienta de simulación Riverbed Modeler [5], anteriormente conocida como Opnet Modeler. Frente a la variada oferta de herramientas de simulación

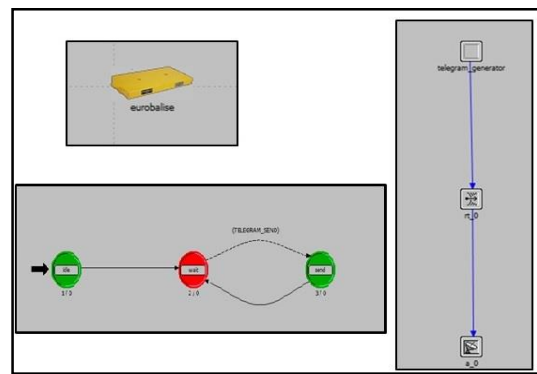


Fig. 3. Diseño de eurobaliza en Riverbed Modeler.

existentes (NS-3, Omnet++, etc.) se ha seleccionado ésta por su amplia gama de librerías que permiten modelar redes en entornos reales y cubren la mayoría de las tecnologías. Además, dentro del grupo de investigación existen proyectos previos realizados con esta herramienta. Sin embargo, ninguno de los modelos proporcionados por este simulador modela comunicaciones de campo cercano como es el caso de las eurobalizas. Por lo tanto, para un correcto modelado de las balizas, ha sido necesario modelar también este tipo de comunicaciones. A continuación se explica en detalle el modelado realizado.

A. Creación de los nodos

El primer paso para este estudio es la creación de los nodos que van a implementar las funcionalidades de eurobaliza y BTM. Los modelos utilizados por la herramienta Riverbed Modeler se basan en nodos que a su vez hacen uso de módulos y procesos, de forma que dichos elementos son los que se han diseñado para la baliza y BTM. La Fig. 3 muestra un esquema de la eurobaliza diseñada en Riverbed Modeler.

La baliza consta de 3 módulos:

- Generador de telegramas
- Transmisor radio
- Antena

El módulo principal es el encargado de la generación de los telegramas. Dicho módulo monitoriza en todo momento el entorno de la baliza para detectar si algún tren entra en el volumen de activación de la baliza, en cuyo caso ésta comienza a emitir información. Para ello, genera el telegrama con la longitud y formato adecuado y lo emite constantemente hasta que no haya ningún tren dentro del volumen de activación de la baliza. Si bien Riverbed Modeler no funciona con flujos de bits sino con paquetes -en este caso telegramas-, es posible enviar un flujo constante de bits ajustando la periodicidad de envío de los telegramas en función de la tasa de transmisión y del tamaño de los telegramas.

Por su parte el nodo de BTM también consta de 3 módulos como se aprecia en la Fig. 4:

- Receptor de telegramas
- Receptor radio
- Antena

El módulo receptor de telegramas de la BTM sólo recibe de los módulos inferiores (receptor de radio y antena) los telegramas que han sido recibidos sin error tras pasar por el

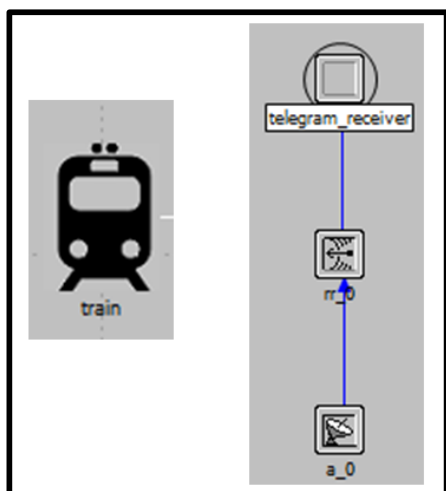


Fig. 4. Diseño de BTM en Riverbed Modeler

canal de transmisión inalámbrico de Riverbed Modeler. Así, es el módulo en el que se puede contabilizar la cantidad de telegramas correctos recibidos.

B. Modelado del canal de transmisión inalámbrico

Para la evaluación del canal radio, los módulos de transmisión y recepción de radio en Riverbed Modeler disponen de un conjunto de etapas denominadas pipelines que se ejecutan de forma secuencial cada vez que se envía un mensaje. Estas etapas se dividen en etapas asociadas a la transmisión del mensaje y etapas asociadas a la recepción del mensaje. Cada una de estas etapas desarrolla una función específica asociada a la transmisión inalámbrica del mensaje, por ejemplo, la comprobación de la correspondencia entre los rangos de frecuencias de transmisión y recepción, el cálculo de la ganancia de las antenas, el cálculo de la relación señal a ruido, etc. La Fig. 5 muestra los pipelines de transmisión asociadas a la eurobaliza diseñada en Riverbed Modeler.

Tanto el módulo de transmisor radio de la baliza como el módulo receptor radio de la BTM utilizan los pipelines estándar de Riverbed Modeler “dra”. Sin embargo, algunas de estas etapas han sido modificadas para adaptarlas al caso específico de las comunicaciones baliza-BTM.

En el módulo transmisor de radio de la baliza, de acuerdo a la Fig. 5, se han modificado los modelos de dos etapas o pipelines:

- Pipeline de closure. Esta etapa estudia si un nodo inalámbrico puede recibir el paquete enviado por un transmisor. Se ha creado un nuevo modelo llamado “eurobalise_closure” que verifica que el receptor se encuentra dentro del volumen de activación de la baliza. En caso negativo descarta al nodo como receptor del telegrama emitido.
- Pipeline de chanmatch. En esta etapa se ha desarrollado un nuevo modelo llamado “eurobalise_chanmatch” que comprueba que el transmisor y el receptor funcionan en el mismo rango de frecuencias, mismo ancho de banda, tasa de bits y modulación. En caso de que algún parámetro no coincida exactamente entre el transmisor y el receptor, el paquete será considerado ruido en el receptor.

Attribute	Value
name	rt_0
channel	(...)
modulation	promoted
rxgroup model	dra_rxgroup
txdel model	dra_txdel
closure model	eurobalise_closure
chanmatch model	eurobalise_chanmatch
tagain model	dra_tagain
propdel model	dra_propdel

Fig. 5. Pipelines en el módulo de transmisión radio de la baliza.

Por otro lado, en el módulo receptor radio de la BTM se han modificado otras etapas del canal de transmisión inalámbrico del simulador:

- Pipeline de Power. Éste es el pipeline encargado de calcular la potencia final con la que es recibido el paquete en el receptor. Las modificaciones realizadas en este apartado son detalladas en la sección III.C.
- Pipeline de BER (Bit Error Rate). En los pipelines anteriores se ha calculado el valor SNR (Signal Noise Ratio) efectivo con el que se ha recibido el paquete. En base a este nivel de SNR efectivo y a la modulación usada se obtiene el BER. En este pipeline no se ha definido un nuevo modelo, se sigue usando el modelo por defecto “dra_ber” para el pipeline, pero ha sido necesario implementar las curvas de modulación CPFSK Coherent (1) y CPFSK Non-Coherent (2) que son las utilizadas en las comunicaciones Baliza-BTM y que no se encuentran implementadas en Riverbed Modeler.

CPFSK Coherent:

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{2N_0}} \right) \quad (1)$$

CPFSK Non-Coherent:

$$P_b = \frac{1}{2} e^{-\frac{E_b}{N_0}} \quad (2)$$

- Pipeline de Error. En este pipeline, una vez conocido el BER, se calculan el número de bits erróneos en el paquete y en función de ese número y del sistema de corrección de errores de la tecnología inalámbrica se marca el paquete como válido o erróneo. El código CRC usado en la BTM es un poco particular porque además de verificar la integridad del telegrama actual también comprueba la integridad de parte del telegrama anterior con lo que ha sido necesario crear un nuevo modelo llamada “eurobalise_error” para esta etapa.

C. Modelado de la transmisión de campo cercano

El modelo “dra_power” usado por defecto en el pipeline de Power por la herramienta de simulación a la hora de calcular la potencia final con la que es recibido un paquete utiliza la ecuación de Friis que tiene en cuenta las pérdidas de potencia debidas a espacio libre. Dicha ecuación relaciona las potencias recibida y transmitida entre 2 antenas separadas una distancia R, que cumple la condición de campo lejano (3).



Fig. 6. Escenario de simulación.

$$R > \frac{2D^2}{\lambda} \quad (3)$$

Sin embargo, las comunicaciones entre la baliza y la BTM son de campo cercano, con lo que no aplica esta fórmula y por lo tanto ha sido necesario crear un nuevo modelo de nombre “eurobalise_power” para el pipeline de Power basado en medidas realizadas en laboratorio.

Así, se han realizado mediciones en laboratorio teniendo en cuenta los siguientes parámetros iniciales: el rango de activación de la baliza va de -1,3 a 1,3 metros a lo largo del eje horizontal, la altura de la BTM respecto a la baliza es de 0,2 metros, la potencia de transmisión de la baliza es de 25 dBm y la sensibilidad de la BTM es de -6 dBm. En base a estos parámetros iniciales se ha ido moviendo la BTM y guardando los valores de potencia recibida por la BTM en cada posición. Conociendo los valores de potencia emitida por la baliza y potencia recibida por la BTM se ha generado una matriz de atenuación en base a la distancia entre baliza y BTM que es la matriz de atenuación que se aplica en el modelo de pipeline “eurobalise_power” en lugar de la fórmula de Friis utilizada en el modelo “dra_power”.

IV. PRUEBAS/RESULTADOS

El principal objetivo de las pruebas es validar el modelo desarrollado en Riverbed Modeler respecto a las medidas reales. Para ello se han evaluado las potencias recibidas por el módulo BTM del tren. Esta relación de potencias varía en función de la distancia entre el tren y la eurobaliza instalada en la vía.

Para desarrollar estas pruebas hemos generado un escenario de simulación básico dotado de una eurobaliza y un tren. Al tren se le ha asignado una trayectoria fija emulando una vía ferroviaria y la eurobaliza ha sido instalada en un punto de esa trayectoria (véase Fig. 6). De esta manera, somos capaces de validar la activación de la eurobaliza cuando el tren la activa mediante su campo inductivo, así como su paso al estado dormido por salir de dicho campo.

Los parámetros utilizados para configurar la eurobaliza quedan reflejados en la Tabla I. Estos parámetros se han establecido partiendo de los requerimientos definidos en [4].

Las pruebas realizadas para analizar el correcto funcionamiento del modelo desarrollado se dividen en tres: la verificación de conectividad, la verificación de las pérdidas de transmisión en campo cercano y por último, el análisis de la tasa de bit erróneos y la relación señal a ruido de las comunicaciones.

Tabla I
PARÁMETROS DE CONFIGURACIÓN EUROBALIZA

Atributo Riverbed Modeler	Value
Data rate (bps)	561.480
Bandwidth (kHz)	565
Min Frequency (MHz)	3,951
Power (W)	0,316
Bit capacity (bits)	Infinity
Pk capacity (pkts)	1.000

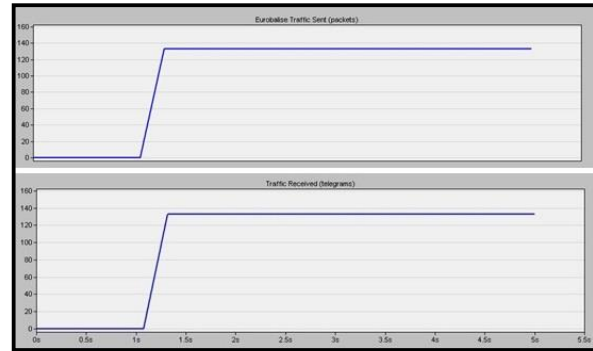


Fig. 7. Telegramas enviados (baliza) y recibidos (tren).

Para la validación de la conectividad entre la eurobaliza y el tren durante el paso del tren sobre la eurobaliza se han recogido los estadísticos referentes a los paquetes enviados por la eurobaliza y recibidos por el tren. En la Fig. 7 ambos estadísticos son correlados validando que los paquetes enviados coinciden con los recibidos y por tanto verificando que la comunicación es correcta.

Una vez verificada la correcta comunicación entre ambos elementos, el siguiente estudio se ha basado en analizar la potencia recibida por la BTM en cada instante temporal cuando el tren se encuentra en movimiento.

Para ese cálculo, se han utilizado los datos de pérdidas obtenidos de mediciones reales [6], los cuales se han incluido en uno de los pipelines que han sido modificados para este modelo (pipeline eurobalise_power), tal y como se ha comentado anteriormente.

Por otro lado, la obtención de la SNR se ha basado en los cálculos obtenidos de la potencia y en los valores de ruido de la transmisión. La Fig. 8 muestra la variación de potencia en el tiempo a medida que el tren se desplaza sobre la eurobaliza. Estas potencias han sido simuladas con el modelo descrito en este artículo. Tal y como puede observarse, este análisis de potencia se corresponde de manera precisa con las muestras reales recogidas en [6] e ilustradas en la Fig. 9.

Por último, se ha evaluado la influencia de las pérdidas en la transmisión. Para ello se han recogido datos sobre la potencia recibida en el tren, la relación señal a ruido (Signal to Noise Ratio, SNR) y la tasa de bits erróneos (Bit Error Rate, BER). Estas medidas sirven para evaluar la influencia que el ruido electromagnético tiene en la señal.

En la Fig. 10 se refleja la evolución de la relación señal a ruido durante el periodo de comunicación de la baliza y el tren.

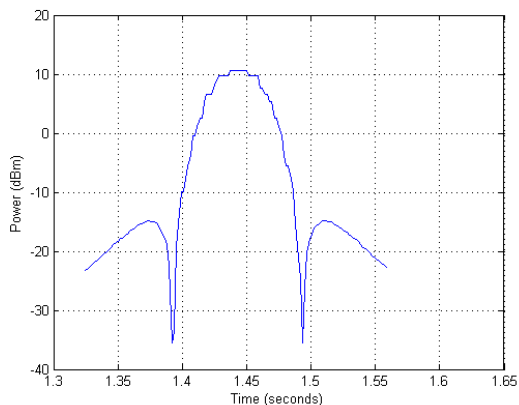


Fig. 8. Potencia recibida en la BTM (Riverbed Modeler).

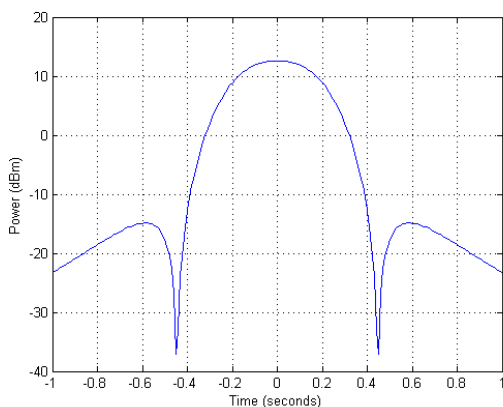


Fig. 9. Potencia recibida en la BTM (muestra real)..

Se observa cómo la relación señal a ruido de la Fig. 10 evoluciona de forma similar a la potencia de recepción expresada en la Fig. 8. Se obtiene un valor máximo en el instante de mayor proximidad entre la baliza y el tren. Por su parte el BER mantiene un valor constante durante toda la simulación. Dicho valor es muy cercano a cero ya que el nivel de SNR es suficientemente alto, de forma que no se producen errores en la transmisión.

V. CONCLUSIONES Y PRÓXIMOS PASOS

En este artículo se describe el modelado en la plataforma de simulación de eventos discretos y propósito general, Riverbed Modeler, de la comunicación entre la eurobaliza y el equipo embarcado en el tren, BTM.

Esta comunicación es una parte del conjunto de comunicaciones tierra tren que abarca el sistema de señalización ferroviario estandarizado para alta velocidad ERTMS. Ahora bien, estas comunicaciones suelen estar sujetas a interferencias y perturbaciones. Un fallo o degradación de los indicadores de calidad de la comunicación eurobaliza-BTM puede afectar significativamente la operación ferroviaria. Al tratarse de aplicaciones críticas de señalización, un fallo en la recepción puede llegar a implicar una parada de emergencia en el tren.

En este artículo se ha presentado una primera etapa en la caracterización de esa comunicación eurobaliza-BTM, en la que se han obtenido indicadores de rendimiento de la capa

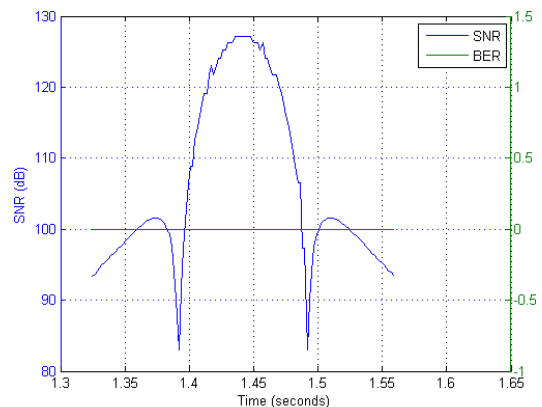


Fig. 10. Relación señal a ruido (SNR) y tasa de bits erróneos (BER).

física como son la potencia recibida el SNR y el BER. Los valores obtenidos para estos indicadores se han validado con medidas realizadas en una maqueta real.

El siguiente paso es caracterizar las interferencias más comunes en este entorno e integrar en nuestro framework de simulación [1], donde se modela el conjunto completo de la aplicación ERTMS. De esta manera será posible establecer la relación entre interferencias en indicadores de calidad de servicio de las capas superiores del sistema ERTMS.

AGRADECIMIENTOS

El trabajo presentado en este artículo se ha realizado en la Unidad de Formación e Investigación UFI11/16 financiada por la UPV/EHU. Este trabajo está apoyado por el Ministerio Español de Economía y Competitividad a través del proyecto de investigación SAREMSIG TEC2013-47012-C2-1-R (Contribution to a Safe Railway Operation: Evaluating the effect of Electromagnetic Disturbances on Railway Control Signalling Systems) y financiado a través del Programa Estatal de Investigación, Desarrollo e Innovación y orientado hacia los Retos de la Sociedad 2013. Este trabajo también se enmarca en el contexto de trabajo del proyecto Europeo EU FP7 Collaborative Research Project titulado SECURITY of Railways against Electromagnetic Attacks-SECRET Project.

REFERENCIAS

- [1] Aguado, M., Pinedo, C., Lopez, I., Ugalde, I., De Las Munecas, C., Rodriguez, L., & Jacob, E. (2014, September). Towards zero on-site testing: Advanced traffic management & control systems simulation framework including communication KPIs and response to failure events. In *Wireless Vehicular Communications (WiVeC)*, 2014 IEEE 6th International Symposium on (pp. 1-2). IEEE.
- [2] Especificaciones ETCS: "Subset-026-2 v3.0.0", ERA, UNISIG, EEIG ERTMS USERS GROUP.
- [3] Especificaciones ETCS: "Subset-116 v1.0.0", ERA, UNISIG, EEIG ERTMS USERS GROUP, pp.21
- [4] Especificaciones ETCS: "Subset-036 v3.0.0", ERA, UNISIG, EEIG ERTMS USERS GROUP.
- [5] Pagina oficial Riverbed Modeler. www.riverbed.com
- [6] L. Zhao, Y. Jiang, "Modeling and Simulation of Balise Up-link data transmission based on finite element method", *Journal of Theoretical and Applied Information Technology*, Vol. 46, No. 2, 2012, pp. 867-874
- [7] I. Adin, J. Mendizabal, J. del Portillo, U. Alvarado, A. Villaro, A. G. Alonso "Design of a BTM Analog Receiver for Unified European Railway Network" XXIII Conference on Design of Circuits and Systems DCIS 08 Grenoble 2008

Adherencia al tratamiento de deformidades craneales con un dispositivo ortopédico sensorizado

Lucero Rosas¹, Julio García², Antoni Oller², Juan Lopez², and Jesus Alcober²

¹FixYou

²UPC/Alteraid

Existe una enfermedad que afecta a niños acabados de nacer que provoca malformaciones craneales y, las posibilidades de tratamiento (quirúrgico) son limitadas hasta que el niño tiene una edad en que no se ponga en riesgo su vida. Este trabajo describe una experiencia basada en la combinación de las disciplinas de ingeniería biomédica y telemática que ayuda a definir un procedimiento de tratamiento para niños entre 4 y 18 meses para el diagnóstico, diseño de un casco, fabricación y seguimiento en forma de registro de adherencia a lo largo de todo el tratamiento.

Index Terms—Deformational plagiocephaly, helmet, empowerment, cloud, adherence.

I. INTRODUCCIÓN

FIXYOU es una start up iniciada en el Instituto Politécnico Nacional de la Ciudad de México que ofrece un tratamiento integral para tratar deformidades craneales en niños entre lo 4 y 18 meses de edad utilizando un casco ortopédico con sensores téxtiles que registra la adherencia de uso y comparte en una nube esta información con las personas interesadas (familia, personal sanitario, etc) en el tratamiento por medio de una aplicación móvil.

La forma de las cabezas de estos bebés con plagiocefalia o braquicefalia [1], tipo de deformidad craneal, no es simétrica lo cual puede desencadenar problemas con el sentido de la visión, oído, problemas con la masticación y algunos problemas de aprendizaje, además de tener efectos psicológicos en el bebé que los padece al encontrarse de aspecto diferente a los demás niños. La manera de corregir estas deformidades es a través de fisioterapia y en casos severos el uso de un casco ortopédico u ortesis craneal.

El uso del casco ortopédico pediátrico se prescribe a bebés entre 4 a 12 meses de edad que presentan aplanamientos craneales. Se utiliza para recuperar la simetría de la cabeza y como consecuencia evitar los problemas que se han descrito anteriormente. Para tener resultados favorables y en el mínimo de tiempo es necesario llevar el casco ortopédico durante 20 horas diarias a lo largo de 6 meses en promedio y por este motivo es importante disponer de un sistema de control de adherencia que nos informe con la máxima fiabilidad y mínima molestia por parte del bebé del uso del casco.

Este proyecto se inició en 2011 con una colaboración que se realizó con el instituto Nacional de Pediatría. Este hospital público es el más importante en México que atiende a pacientes, que provienen de todo el país, de las características descritas anteriormente. El proyecto ha sido supervisado por dos neurocirujanos pediatras y el apoyo de las instalaciones del hospital.

Este trabajo describirá este sistema de adherencia integrado en el casco ortopédico y se organizará de la siguiente manera.

En el segundo apartado se describirá el estado del arte de los sistemas de adherencia. En el tercer apartado se enumerarán los retos y objetivos del proyecto. En el cuarto apartado se presentará la arquitectura completa del sistema y, finalmente se analizarán los resultados y conclusiones de este trabajo.

II. ESTADO DEL ARTE DE LOS SISTEMAS DE ADHERENCIA A LOS TRATAMIENTOS

Los sistemas de adherencia al tratamiento que se han desarrollado hasta la fecha generalmente se enfocan a los del tipo farmacológico. Los métodos utilizados se clasifican en diversos tipos:

- *Técnicos*: cómo simplificar las dosis a una sola toma diaria o paquetes simplificados para dosis diarias. Sistemas de *seguimiento electrónico* que en general son alarmas y sensores que cuentan la cantidad de medicamento restante.
- De *Comportamiento*. Consisten en recordatorios por vía telefónica, aplicaciones móviles [5], mailing, visitas al hogar de los pacientes, etc. También se ha trabajado con sistemas a base de recompensas (gamificación) en este tipo de método.
- *Educacional*: consiste en ofrecer información para que el paciente tenga un mayor entendimiento de su enfermedad, como información para diabéticos o con hipertensión. Se encuentra también como métodos de adherencia al tratamiento la asistencia a grupos de apoyo.
- *Multifacético*: Se combinan dos o varios métodos. En la mayoría de los artículos revisados los investigadores alertan sobre la falta de algún protocolo método fijo sobre herramientas de adherencia tanto a tratamientos farmacológicos como no farmacológicos. [7] [8]

III. RETOS Y OBJETIVOS DEL CASCO SENSORIZADO

La edad clave para tratar este tipo de deformidades (ver figura 1) es entre los 4 y 18 meses de edad, ya que después el cráneo se consolida y ya no es posible realizar correcciones. Cada casco se hace a la medida de paciente y puede ser de plástico o reforzado con fibra de vidrio en el exterior. En el

interior está recubierto de goma ortopédica. Debe usarse la mayor parte del día durante 6 meses en promedio. Para darle seguimiento cada 5 semanas se cita al paciente para observar la evolución y ajustar el casco, pues la cabeza del bebé continúa con su crecimiento natural.

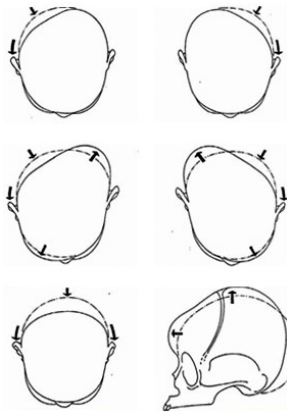


Fig. 1: Malformaciones del cráneo

Conforme ha evolucionado el proyecto se han encontrado varios problemas a resolver. Inicialmente lo que se busca, es que todos los bebés tengan acceso a este tratamiento, pues para el caso de México no es posible encontrar centros donde les den atención, a diferencia de países como España, EEUU, Reino Unido o Francia por nombrar algunos. Esta situación, obliga a los padres a trasladarse a Estados Unidos, o en el peor de los casos, dejar el padecimiento de su bebé sin atender. Los ingenieros en FixYou aplicando tecnología e ingeniería han trabajado sobre un producto que las personas del tipo de economías como México y países de Latinoamérica puedan costear pero que al mismo tiempo tengan la seguridad que están usando de calidad que cuidará la salud de sus bebés.

El siguiente problema que se identificó, ya con varios pacientes usando el casco, es la adherencia al tratamiento que es clave para lograr una corrección. Actualmente no es posible conocer la cantidad de horas de uso del dispositivo y hay que confiar que la información que proporciona la familia sobre la adherencia del casco en el bebé, es la correcta (ya sea por falta de memoria o por falsear la realidad). Por este motivo se está trabajando sobre un sensor de uso, ubicado en lugares estratégicos, que sea capaz de definir si el dispositivo está bien colocado o no y el tiempo en el que está en funcionamiento. Para lograr este cometido, se ha realizado un trabajo en conjunto con la empresa AlterAid introduciendo la idea del contador de tiempo de uso, la implementación de una aplicación móvil y sincronización de datos con la nube. Este escenario permitirá a todos los interesados en el tratamiento del bebé, dar seguimiento al uso del casco desde cualquier lugar.

Este sistema de tiempo de uso y adherencia al tratamiento que se ha pensado, puede aplicarse en muchos más casos de medicina, como puede ser para el caso de uso de férulas u otros dispositivos ortopédicos y, resultará de gran utilidad

para verificar resultados de ciertos protocolos de investigación médica.

Por otro lado, el uso de tela e hilos conductivos ha cobrado popularidad en los últimos años por el auge de los dispositivos wearables. Sin embargo, son pocos los trabajos que relacionan el uso de los textiles conductivos con algún sistema de adherencia o registro del tratamiento.

Finalmente se pretende, como objetivo, dar difusión sobre el tratamiento y prevención de este tipo de deformidades, pues se ha encontrado que a nivel global no existe un protocolo para dar atención a estos casos.

IV. ARQUITECTURA DEL SISTEMA

Una familia que tenga un bebé con malformación craneal y que participa en las pruebas piloto de fixyou utiliza un casco ortopédico sensorizado durante varias horas al día. Por otro lado, todas aquellas personas involucradas en el bienestar de este bebé (familia, personal sanitario, etc) puede recibir información de adherencia que proviene del caso y ha sido transmitida a una nube que almacena datos de la salud del bebé. Esta plataforma permite compartir esta información con todos aquellos usuarios que esten interesados en el bienestar del bebé.

El sistema propuesto integra sensores de presión dentro del casco ortopédico. La información recogida por estos sensores es procesada dentro un microcontrolador (basado en tecnología de arduino) montado en forma de una pequeña caja al dispositivo ortopédico para posteriormente ser transmitida a un dispositivo móvil vía bluetooth (BluetoothLE). Dicho móvil será el encargado de enviar esta información a la nube donde estará disponible para todos los interesados en el tratamiento del paciente; el paciente mismo, familiares, cuidadores y médicos tratantes, generando así un red donde los involucrados en el tratamiento del paciente intercambien información y contenido social sin importar el lugar donde se encuentren.

Debido a que este sistema está previsto que sea utilizado por pacientes, su uso debe ser sencillo. Por este motivo, durante el desarrollo de cada etapa se toman en cuenta la complejidad técnica, la usabilidad y experiencia del usuario final.

La siguiente figura (2) muestra los principales componentes y actores que interactúan en el ecosistema propuesto.

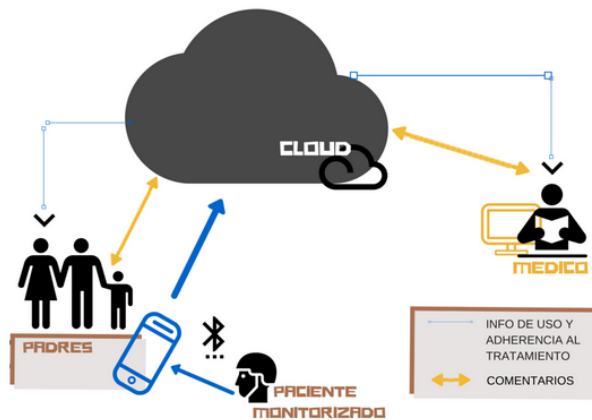


Fig. 2: Arquitectura general del sistema

Para entender con más detalle, el sistema del registro de adherencia al uso del dispositivo ortopédico se deben tener en consideración los siguientes apartados:

- Sensores
- Procesamiento y tratamiento de la información
- Análisis y difusión de información

A. Sensores

Los sensores propuestos para el sistema son de presión a base de tela e hilos conductivos (Ripstop). Se han elegido este tipo pues estarán en contacto directo con la piel y su sudoración característica. Así mismo se cuenta con el factor de que entrarán en contacto con agua al momento de hacer la limpieza del dispositivo. Esta medida evitará su desgaste y funcionamiento incorrecto. La colocación de los sensores tiene posiciones estratégicas (ver figura 3) que además de aportar información para realizar la cuenta de horas de uso y adherencia al tratamiento, también indican si el dispositivo está bien colocado. Ambos sensores están integrados en el interior del casco y para el usuario son prácticamente imperceptibles.

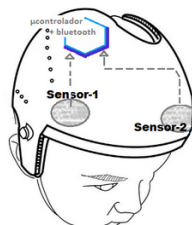


Fig. 3: El casco

B. Procesamiento y tratamiento de la información

Una vez acoplados los sensores, las señales de éstos entran en modo digital a un microcontrolador. Para el trabajo se ha elegido un AVR y para las pruebas preliminares se ha elegido la plataforma de desarrollo Arduino Uno.

Inicialmente el microcontrolador se encuentra en mínimo consumo de energía esperando a la activación a través de rutina de interrupción de alguno de los sensores, posterior a la activación del primer sensor pregunta por el segundo con el fin de identificar si el dispositivo ortopédico se encuentra colocado de manera correcta. En cuanto ambos la información de ambos sensores es positiva se inicia el cuento de tiempo de uso en minutos. Mientras los sensores están activos y el micro se encuentra contando el tiempo éste se enciende y apaga de manera intermitente, con el fin de optimizar la fuente de alimentación del chip, hasta que deja de percibir alguno de los dos sensores. El valor del tiempo transcurrido se almacena en un registro de memoria no volátil. Para realizar la transmisión del valor del tiempo de uso es necesario activarlo desde un switch. La transmisión de la información se hace usando un módulo bluetooth HC-05 que se sincroniza con la aplicación móvil. La aplicación móvil se ha implementado sobre el sistema operativo android y actualmente se trata de una prueba de concepto que permite vincularse al dispositivo de bluetooth, recibir la información de adherencia del casco (un contador de tiempo que representa el tiempo en el que el casco ha estado utilizándose) y la identificación del casco. Una vez recibida esta información, y si se dispone de conectividad, se enviará esta información al repositorio de datos (nube). En caso que el móvil no disponga de conectividad se almacenará esta información en una base de datos local orientada a objetos (db4o). En el momento que exista conectividad, la aplicación móvil enviará toda la información utilizando HTTP seguro (https) y un token de acceso, que previamente se ha conseguido mediante la negociación OAUTH 2.0, que servirá de mecanismo de autorización sobre la API del repositorio de datos. De esta manera se autoriza el envío de información a la nube y se impide que los usuarios puedan falsear la información de adherencia de un casco, ya que todo el proceso es transparente a ellos. Por otro lado, las familias de los bebés con malformaciones están muy implicados en el bienestar de sus bebés y no es el perfil de usuarios que falsean la información, más bien necesitan herramientas de soporte para gestionar información y evitar errores humanos de memoria. La siguiente figura (4) muestra un diagrama de flujo del funcionamiento descrito anteriormente.

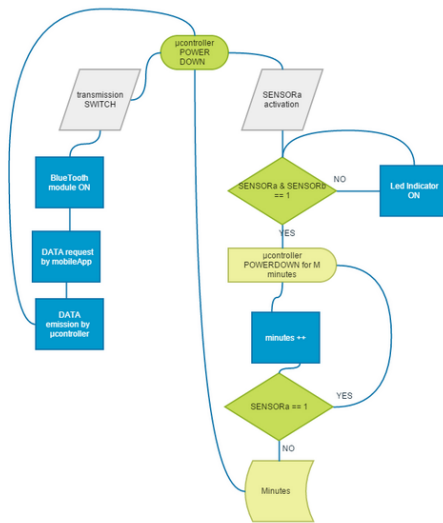


Fig. 4: Diagrama de flujo

Actualmente el proyecto está en una fase piloto y el microcontrolador (basado en tecnología arduino) y los sensores pueden estar alimentados por pilas AA (que pueden ser sustituidas), o una batería LiPo que puede ser recargada cuando sea necesario. El microcontrolador puede monitorizar el estado de la batería y realizar avisos por medio de un led luminoso y/o notificaciones a la aplicación móvil. De esta manera se puede asegurar que el sistema de adherencia al casco pueda funcionar a lo largo de su tratamiento (6 meses aproximadamente). Por otro lado, la información de adherencia del casco se guarda en la memoria no volátil del microcontrolador, por lo que si la batería o las pilas se agotan no se pierde esta información.

C. Análisis y difusión de información

En la arquitectura, la nube (cloud) es la parte central del sistema y se define como SaaS (Software as a Service). Éste es un componente importante porque permite procesar la información adquirida a través de diferentes métodos de entrada (sensores, aplicaciones móviles, etc.).

Para conseguir el acceso a estos datos procesados, este módulo proporciona una interfaz de aplicación (API) basado en REST utilizado para la adquisición de datos y su consumo. Esta plataforma se ha diseñado para ofrecer acceso a terceros, es decir, las aplicaciones, los servicios o los sensores a través de un Interfaz de programación de aplicaciones (API) para la salida y entrada de información. Para garantizar los mecanismos de autorización de acceso sobre la API se utiliza un mecanismo de autorización basado en OAuth 2.0.

D. Modelo de datos

La información del ecosistema social almacenada en la nube puede mostrarse a los usuarios interesados a través de un navegador o con una aplicación móvil. Típicamente se identifican tres tipos de actores diferentes, aunque podrían existir otros en un futuro:

- usuarios familiares: Son usuarios vinculados a un bebé con una malformación craneal. El bebé no será un usuario

en este sistema aunque en un futuro, cuando se haga mayor, podría serlo y visualizar toda la información relacionada con él.

- personal sanitario: El personal sanitario podría desear tener acceso a la información de los pacientes y de esta manera tener un buen seguimiento de sus tratamientos. Al mismo tiempo puede obtener información agregada sobre todos los pacientes.
- Otros miembros de la comunidad: Personas con alguna relación con el bebé. Por ejemplo cuidadores,

La información está modelada y definida en una estructura de datos basada en grafos. Los nodos de este grafo representan un tipo de información que está relacionada entre ella según unos criterios de jerarquía y relación de compartición. Cada nodo almacena información en un paradigma schemaless que permite que existan diferentes conjuntos de campos para cada nodo. Este tipo de organización de la información permite ser flexible y adaptarse a nuevos requerimientos. Este esquema proporciona versatilidad en la organización y ayuda a organizar la información de una manera eficiente.

Las principales entidades de información del modelo de datos son las siguientes:

- Usuarios: Se trata de uno de los componentes principales de la plataforma que gestiona la información relevante de una persona física como el nombre, información de contacto (teléfonos, correo electrónico) para las notificaciones y credenciales de acceso a la plataforma.
- Vínculo: Es un elemento virtual que permite almacenar información o parte de información que se almacenará y compartirá a otros usuarios.
- Medidas: Es el elemento sobre el que se almacenan los valores. La configuración de las medidas es especificada en el momento de su creación. Se puede determinar el tipo de datos (numérico, cadena de caracteres, etc) y metainformación como rangos de valores para lanzar notificaciones, alarmas o análisis de minería de datos (bigdata).

Tal como se ha descrito anteriormente, la naturaleza de la información es jerárquica y por este motivo y después de un análisis de diferentes opciones de bases de datos se ha determinado almacenar la información en una base de datos NOSQL (Not-only SQL) basada en grafos (Neo4j) que está especialmente pensada en almacenar modelos basados en grafos y grandes volúmenes de información.

En este proyecto, la organización de la información se ha definido de la siguiente manera. El bebé (y su información asociada al casco) será un vínculo cuyo propietario será un usuario (por ejemplo, un progenitor). El propietario del vínculo será el que decidirá las políticas de compartición de la información o parte de ella almacenada en este elemento virtual. Por ejemplo, este vínculo que representa y almacena información de un casco de un determinado bebé, puede estar relacionado con diferentes usuarios: familiares, los dos médicos neurocirujanos pediatras, enfermeras pediátricas y la cuidadora del bebé.

A efectos de la estructura de datos existirá un nodo de tipo usuario (el propietario o controlador de la información)

que estará relacionado con nodos de tipo vínculo y uno de ellos será su bebé. Este vínculo tendrá una serie de medidas y una de ellas puede gestionar la adherencia al uso del casco. Adicionalmente se pueden almacenar otro tipo de medidas como las medidas del tamaño craneal, el peso del bebé, etc. Esta información puede ser compartida con otros usuarios y esto implicará que el grafo se enriquezca con nuevas relaciones entre los nuevos usuarios y los vínculos.

V. RESULTADOS

Actualmente se ha logrado testear el dispositivo en 60 pacientes diagnosticados con plagiocefalia sin el módulo de adherencia incorporado. En el futuro estudio se realizará unas pruebas con nuevos pacientes con el sistema de adherencia (módulo de comunicaciones y aplicación móvil) y se podrá evaluar su efectividad del tratamiento en ambos grupos.

A continuación se presentan los resultados de dos pacientes muestra, Px1 y Px2, el tratamiento se llevó a cabo con el acompañamiento de un médico pediatra Neurocirujano.

Para el Paciente 1 [Px1], el tratamiento se inició a los 13 meses de edad, se recogieron datos cada 30 días sumando un total de 6 citas. El Paciente 2 [Px2] fue diagnosticado a los 6 meses de edad pero no se inició el tratamiento hasta los 8 meses de edad, de este paciente se presentan 3 citas de seguimiento.

Las tablas siguientes muestran la evolución del índice oblicuo craneal calculado, este índice indica el grado de asimetría de la cabeza del paciente (Px1). De acuerdo a literatura médica un índice de 104% o menor corresponde a una cabeza simétrica o ‘normal’.

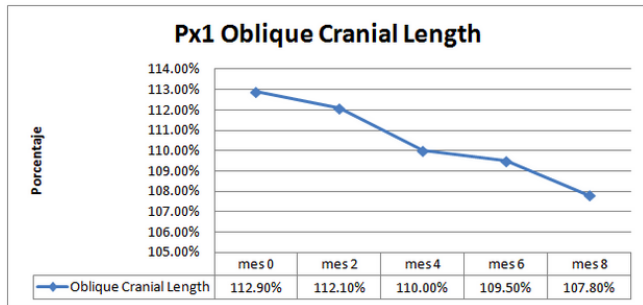


Fig. 5: Resultados paciente muestra Px1

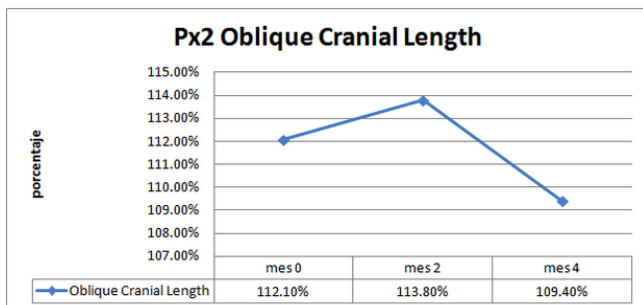


Fig. 6: Resultados paciente muestra Px2

De los gráficos se observa que ambos pacientes la deformidad o asimetría tiende a disminuir y acercarse al valor de 104% correspondiente a una cabeza simétrica. Para el Px2 en la segunda cita se observa una cresta. Este dato corresponde con el hecho de que al ser diagnosticado transcurrieron dos meses sin que el paciente recibiera tratamiento alguno, por lo que la asimetría aumentó, sin embargo al iniciarlo se observa que el valor del índice craneal empieza a disminuir.

A. Resultados esperados

Al tener los datos del tiempo se podrá confirmar la hipótesis de que a mayor tiempo de uso, mayor y más rápida será la corrección, esto se verá reflejado en el valor de la pendiente de la recta correspondiente al índice craneal.

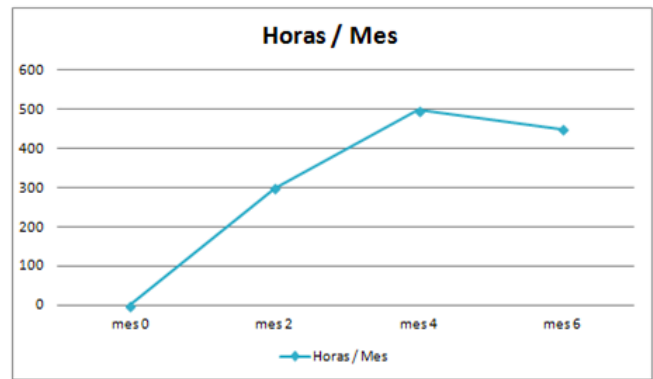


Fig. 7: Adherencia de uso del casco

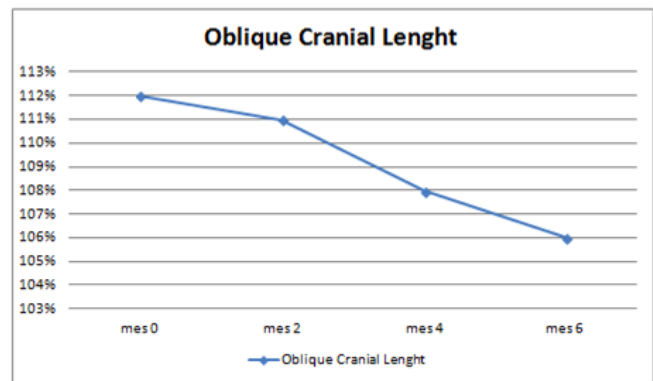


Fig. 8: Índice craneal

VI. CONCLUSION

Hasta el día de hoy se han visto 60 pacientes de entre 4 y 18 meses de edad, los cuáles han usado casco y la mayoría de ellos han tenido resultados exitosos, entendido por exitoso que la forma de su cabeza ha recuperado su simetría. La siguiente figura (9) muestra ejemplos reales de uso.



Fig. 9: cascos reales

La ventaja de este modelo propuesto es que los sensores son imperceptibles para el paciente, y la adquisición de esta información no queda en manos del usuario, si no que se hará automáticamente. Para el caso la transmisión de datos del dispositivo ortopédico al móvil debe involucrarse a la persona (el bebé no podría hacerlo), Por otro lado, el microcontrolador tiene una memoria de respaldo y a través de una aplicación móvil se enviarán recordatorios para la sincronización.

FixYou ha tenido la oportunidad de trabajar con diversas instituciones y organizaciones. FUMEC México-USA; Idea Foundry in Carnegie Melon University, Pennsylvania-USA, Fundación Televisa, México. Y cómo se ha comentado actualmente trabajamos con la empresa AlterAid. Hemos tenido la fortuna de ser reconocidos en los siguientes certámenes:

- Premio Iberoamericanos al Innovación y el Empredimiento by Secretaria General Iberoamerica (SEGIB), Spain 2013.
- Fundación Educación Superior empresa (FESE), México 2014
- Premio a la Innovación Empresarial- Banco Santander SA, México 2014.

REFERENCES

- [1] Trip Database Limited, *What is the role of helmet therapy in positional plagiocephaly?* , TRIP Database Limited, 2009.
- [2] DeGrazia, M; Giambanco, D; Hamn, G; Ditzel, A; Tucker, L; Gauvreau, K *Prevention of Deformational Plagiocephaly in Hospitalized Infants Using a New Orthotic Device* Jognn-Journal of Obstetric Gynecologic and Neonatal Nursing. Volume: 44 Issue: 1 Pages: 28-41 DOI: 10.1111/1552-6909.12523. Published: JAN-FEB 2015
- [3] Kim,SY; Park, MS; Yang, JI; Yim,*Comparison of Helmet Therapy and Counter Positioning for Deformational Plagiocephaly*. SY. Annals of Rehabilitation Medicine. Volume: 37 Issue: 6 Pages: 785-795 . Published: 2013
- [4] Lee, A; Van Pelt, AE; Kane, AA; Pilgram, TK; Govier, DP *Comparison of perceptions and treatment practices between neurosurgeons and plastic surgeons for infants with deformational plagiocephaly* Neurosurg Pediatr. 2010 Apr;5(4):368-74. doi: 10.3171/2009.11.PEDS0983.
- [5] L. Dayer, S. Heldenbrand, P. Anderson, P. O. Gubbins, B. C. Martin *Medication Adherence Apps. Potential Benefits to Patients and Providers* Journal of the American Pharmacists Association, 2013;53(2):172-181.
- [6] Rogers, GF; Miller, J; Mulliken, JB. *Comparison of a modifiable cranial cup versus repositioning and cervical stretching for the early correction of deformational posterior plagiocephaly* Plastic and Reconstructive Surgery. Volume: 121 Issue: 3 Pages: 941-947. DOI: 10.1097/01.prs.0000299938.00229.3e. Published: MAR 2008.
- [7] S; Sluijs, E; Van Dijk,L; De Ridder, D; Heerdink, R; Bensing, J *Patient adherence to medical treatment: a review of reviews Van Dulmen* BMC Health Services Research 2007, 7:55
- [8] Hall, AM; Kamper, SJ; Hernon, M; Hughes, K; Kelly, G; Lonsdale, C; Hurley, DA; Ostelo, R *Measurement Tools for Adherence to Non-Pharmacologic Self-Management Treatment for Chronic Musculoskeletal Conditions: A Systematic Review* Archives of Physical Medicine and Rehabilitation. Volume: 96. Issue: 3. Pages: 552-562. DOI: 10.1016/j.apmr.2014.07.405. Published: MAR 2015

Influence of the path establishment method on an SDN controller time response

David José Quiroz Martiña, Cristina Cervelló-Pastor
Department of Network Engineering,
Universitat Politècnica de Catalunya,
Esteve Terradas, 7, 08860, Castelldefels, Spain.
david.quiroz@estudiant.upc.edu, cristina@entel.upc.edu

Abstract—This paper aims for the assessment of a layer 3 path establishment method, to explore the time response added to an Software Defined Networking (SDN) controller during network events and some additional operations, measured from an SDN path methodology used as a reference. The experience learned in this experimentation is presented as an alternative to monitor the processing time that new path establishment methods can bring to an SDN environment.

Index Terms—Software Defined Networking (SDN), controllers, control layer.

I. INTRODUCTION

Software Defined Networking (SDN) has acquired great acceptance in the networking community. Thus, several developments and advances have been made to introduce new functions and applications that can give more maturity and specialized operations to the system. Some developments include the adaptation of layer 3 (L3) protocols, such as IGP and BGP to support routing methods into the SDN environment. Since the principle of SDN is a centralized control of the network, every operation triggered by these protocols implies processing introduced to the controller and a response time, that may or may not have an impact over the network.

Keeping track of the response time during the adaptation of a new path method to SDN, can help developers setting up a margin, to fulfill a balance between the time required to address the processing needs of the path method, and maintaining a lower response time from the controller.

One example of a L3 path establishment method that is being adapted to SDN is Segment Routing. This method constructs an end to end path based on a sequence of nodes and port codes called segments [1]. On the first node, a packet is guided by this sequence of segments. Once the packet passes through each segment, its related segment id (SID) is pulled out of the sequence. This behaviour is maintained until the packet arrives to the edge node that connects to the destination.

This article conducts a series of measurements that helps to determine the approximated response time that Segment Routing can introduce into an SDN controller during normal operations and during network events. The response times are compared with the performance of the Proactive Forwarding method. This mechanism is a L2 path establishment procedure, traditionally used in SDN to construct paths according to traffic destination addresses, and set up by flows installed on

all switches relevant to the path. The response times of this method are used as a reference to visualize the additional time response introduced by Segment Routing, with respect of controller's default state where Proactive Forwarding is supported.

Both path methods are established by the controller into the network using the OpenFlow protocol as their interface between the control plane and data plane. The actions triggered by both methods are exchanged between the controller and the network devices through OpenFlow messages.

OpenFlow has been the first protocol used in SDN as the communication interface between the control plane and the forwarding plane, found in network devices such as switches and routers. It is widely used in most of the current SDN controllers to address the dynamic control of network resources according to the needs of today's applications [2]. It uses TCP sessions to carry out instructions given by the SDN controller, and programs the flow tables found in different network devices, to set an OpenFlow pipeline where the forwarding process is carried out by entries found in the flow tables.

This paper will show that, according to the method applied, the actions triggered by both path methods will determine the way of how the OpenFlow pipeline will be used on the network devices during packet forwarding.

The remainder of this paper is organized as follows. Section II describes the network scenario, including the test scenarios and the elements specifications. Section III illustrates the testing process of each path mechanism and their results. Section IV is devoted to the analysis of the experimental results obtained in the previous section. Finally, Section V presents the conclusions and future work.

II. NETWORK SCENARIO DESCRIPTION

The general idea to measure the behavior of path methods in front of network events is to provide a scenario where an end to end communication is available through multiple paths. A virtual network topology was built to meet these conditions, and is presented on Fig. 1. The network topology is comprised of 10 L2 virtual switches interconnected with an SDN controller, using a TCP port per switch-controller connection over a single ethernet link of 1Gbps, 2 virtual hosts (h1 and h2) from where the end to end communication is taken place, and 12 virtual Ethernet links interconnecting

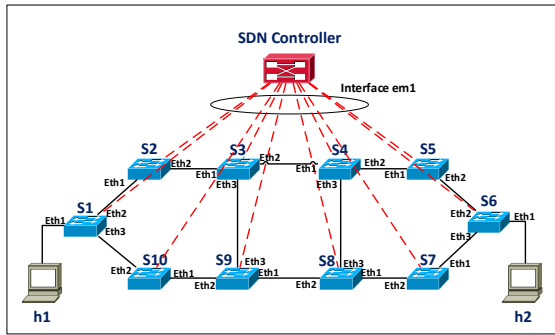


Fig. 1. Test Network Topology.

the switches. The transmission queues of these switches have a default length of 1000, and their bandwidth are dependent of the limitations of the virtual switch. The links disposition allows the SDN controller for the computation of 8 possible paths between hosts h1 and h2 as listed below.

- 1) S1, S2, S3, S4, S5, S6
- 2) S1, S10, S9, S8, S7, S6
- 3) S1, S2, S3, S9, S8, S7, S6
- 4) S1, S2, S3, S4, S8, S7, S6
- 5) S1, S10, S9, S3, S4, S5, S6
- 6) S1, S10, S9, S8, S4, S5, S6
- 7) S1, S2, S3, S9, S8, S4, S5, S6
- 8) S1, S10, S9, S3, S4, S8, S7, S6

The elements involved to build this scenario are comprised of 2 physical computers, one to host the SDN controller, and the other one to host the virtual network topology, both meeting the minimum requirements to run the respective systems. The network topology is built using the Mininet software with CPqD user space switch running OpenFlow 1.3. The SDN controller is an Open Network Operating System (ONOS) controller configured to use the described path establishment methods [3].

A. Test scenarios

There are four test scenarios that were used to evaluate the performance of Proactive Forwarding and Segment Routing path methods: 1) Network performance, 2) Response time in front of network events, 3) Static Path installation time and, 4) Switch forwarding delay.

1) *Network performance*: It is measured under steady state conditions, meaning that the network topology does not experience any network event that can change the network state during the test. Moreover, the test is performed after the network is stabilized on its initial traffic forwarding, in which the controller is consulted for forwarding decisions and paths are installed in the process, adding delay to the initial traffic. The goal is to observe how the network topology behaves in terms of traffic forwarding without the influence of external factors allowing to identify the limits of the topology, and to ensure proper testing using stable values of bitrates. A

traffic is transmitted from one host to the other, using traffic generators that allows to send traffic in different bitrates using UDP datagrams.

Traffic is measured in five main bandwidths: 100 Kbps, 1Mbps, 10Mbps, 100 Mbps, and 1 Gbps, using UDP datagrams of 1400 Bytes. The results to observe are the receiving bitrate and packet loss percentage. Using the more stable bandwidth (0% packet loss and end to end sustained bitrate), the round trip time and jitter are also measured.

2) *The Response time*: This time is measured under network event conditions, where a constant rate traffic is generated from one host to the other, and a network event such as link failures occur during transmission. The link failures are emulated by shutting down specific links of the network in the Mininet console while the traffic is being forwarded through those links.

The Wireshark tool is used to measure the time that takes the controller to reroute the paths and redirect the traffic. OpenFlow messages exchange between the SDN controller and the switches are captured based on [3]. The time frame is measured between the instant where the network event is detected by the SDN controller (OFPT_PORT_STATUS message), and the last flow message (OFPT_FLOW_MOD) sent by the SDN controller. This is the time period in which the controller starts and finalize the process of path redirection, and hence, the response time in front of network events. The processing time of the switches to detect the failure, and send the port status messages are not taken into account. Thus, the measurement only focuses on the performance of the SDN controller to react in front of network events. Throughout this test, the Wireshark tool is also used to constantly monitor the traffic coming through the network, in order to identify the correct link to emulate the failure (more used in Segment Routing compared with Proactive Forwarding, where the traffic can be monitored through the controller's Graphic User Interface).

The Wireshark tool is installed in both equipments (the computers hosting the virtual network and the SDN controller) to perform the described measurements. For the capture of OpenFlow messages Wireshark probes the SDN controller's interface with the switches (interface em1 at Fig. 1). Moreover, for the capture of UDP traffic transmitted by host h1, Wireshark probes the links S3-S4 and S8-S9 through the corresponding interfaces (S3-eth2 and S9-eth1). No matter what path is calculated by the controller, the traffic will pass through either one or both links in all cases.

3) *Static path installation*: Static paths are those in which they can be manually programmed from the Command Line Interface (CLI), and they are measured under steady state conditions. The time that takes between the command execution on the CLI and the last OFPT_Flow_Mod message sent by the controller, is the controller's time response to program and install a static path in the network. This is done in order to observe the effects of the controller response during a manual path installation.

The Wireshark tool probes the network in the same disposition as described in the previous section. The only difference is

that in the computer hosting the SDN controller, the Wireshark probes not only the interface connecting to the virtual network (em1), but also to a secondary interface where it can capture instructions sent by a remote computer (wlan0), as illustrated on Fig. 2. This remote computer is used to initiate a remote CLI session with the controller, and send the execution of the path installation. Wireshark can capture the execution timestamp, and compare it against the subsequent OpenFlow message time stamps, allowing to measure the time frame under a common time reference.

4) *Switch packet forwarding delay*: This time is measured under steady state conditions. The delay observed for an IP packet to be forwarded by a switch in the network corresponds to the time that takes this IP packet between entering an input switch port, and forwarded to an output switch port. This is done in order to observe the effects of how the path methods use the OpenFlow Pipeline of the switches. In this case, the wireshark tool is used on switch S2, chosen randomly for this test to probe its interfaces S2-eth1 as the input port, and S2-eth2 as the output port.

B. Elements specifications

The specifications of the involved elements in the network scenario are the following:

- *Computer hosting the SDN controller*
 - 3.4GHz Intel(R) Core(TM) i7-3770 processor
 - 16GB RAM memory
 - 64bit Ubuntu 14.04.01
- *Computer hosting the network topology*
 - 3.4GHz Intel(R) Pentium(R) D dual core processor
 - 1GB RAM memory
 - 64bit Ubuntu 14.04.01
- *ONOS Controller versions*
 - Blackbird release 1.1.0rc2
 - Spring-Open
- *Mininet release 2.2.0*
- *CPqD Software switch release 1.3.0*
 - User-space switch
 - OpenFlow 1.3
- *Measurement tools*
 - Wireshark version 1.10.6
 - Iperf version 2.0.5

III. EVALUATION OF THE PATH METHODS

Both Proactive Forwarding and Segment Routing mechanisms have demonstrated their capacity to sustain traffic during network events and manual configuration of paths, despite of their differences in path processing times. The following subsections illustrate the testing process of each path mechanism and their results.

A. Proactive Forwarding Testing

Using the network topology, Proactive Forwarding is applied using the intent forwarding application of the Blackbird version ONOS controller [4]. Since Proactive Forwarding is

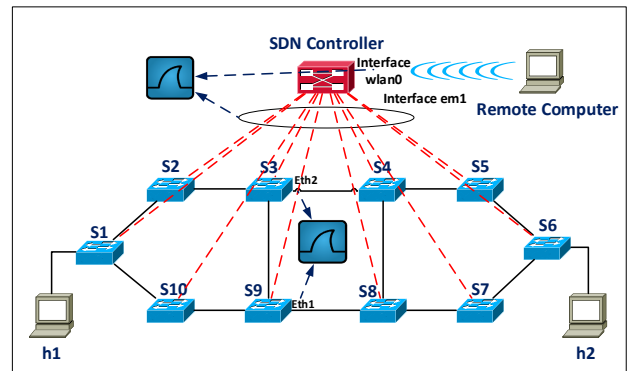


Fig. 2. Wireshark Probe Location.

used as a layer 2 routing mechanism, the virtual hosts (h1 and h2) IP addresses are configured under the same subnet as shown in Fig. 3.

For the network performance test, Iperf was used to generate traffic comprised of UDP datagrams within IP packets, using a constant bitrate, and measuring on intervals of 1 second during 1 minute test. The average results observed in the test are illustrated on Table I.

Up to traffic of 100 Mbps there is a huge packet loss detected on the network, and on the rates of 1 Mbps and 10 Mbps, despite of not presenting packet losses, the bandwidth is not maintained throughout the network, which means that the subsequent tests have to be made under 1 Mbps rates. This is a limitation introduced by the virtual switch by not working at kernel space, which cannot take full advantage of the hardware to sustain higher bandwidth.

Using the rate of 100 Kbps the values of jitter and round trip time (RTT) are measured to keep track of the initial conditions of the network before the subsequent tests. The average round trip time observed was 1.54 ms and an average jitter of 0.043 ms.

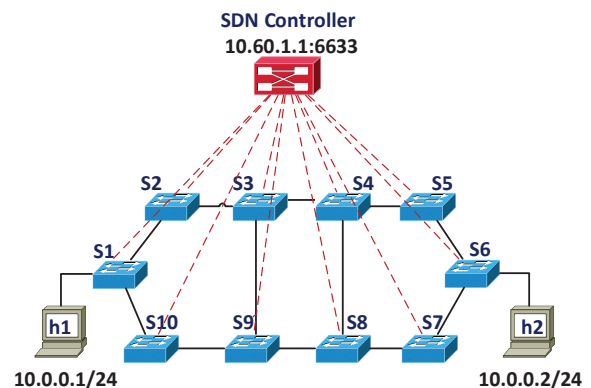


Fig. 3. Proactive Forwarding Topology.

TABLE I
NETWORK PERFORMANCE USING PROACTIVE FORWARDING.

Bitrate	Packet Loss	Received Bitrate
100 Kbps	0%	100 Kbps
1 Mbps	0%	412 Kbps
10 Mbps	0%	2.24 Mbps
100 Mbps	78%	3.31 Mbps
1 Gbps	96%	2.89 Mbps

For the response time testing, Iperf generates a constant rate traffic, measuring on intervals of 1 second during a time frame of 20 seconds in which the link failure is emulated. A number of 100 measurements were made for the network response, executing failures in specific links where the traffic was passing through. Since the path computation is dynamic, not always the SDN controller computed the same initial path on the network, but in most cases it resolved the shortest paths (1 or 2 as described in section II). The emulated failures were links S3-S4 and S8-S9, since the traffic will always pass through either one or both links in every possible path. The controller always recalculates the shortest path after a failure, which in most cases were also paths 1 or 2 described in section II. Table II shows the minimum, maximum, and average time of response that the controller took to process the path redirection, and to completely install it into the network. It also shows the sample standard deviation (σ) and the 95% Confidence Interval (CI).

A time frame is measured between the first port status message received by the controller, and the first OFPT_FLOW_MOD message sent by the controller to the switches. This is the approximated time that the controller takes to process the information received about the change occurred in the network topology [3]. The time between the first port status message and the last OFPT_FLOW_MOD message, is the overall time that takes the controller to process the topology change information, and completely install all the necessary flows to redirect the traffic to the new path [3]. In all the measurements made, there was a minimum traffic interruption observed (0.0012% of average packet loss) between hosts h1 and h2, with a 95% confidence that the mean jitter maintains on the range between 0.0592 ms and 0.0709 ms.

In other observations made during the test of Proactive Forwarding, the paths created not always were multiple paths. In other words, the path computed by the controller, in many

TABLE II
RESPONSE TIME USING PROACTIVE FORWARDING.

Description	Until First Flow_Mod	Until Last Flow_Mod
Minimum	22.513 ms	28.554 ms
Maximum	55.978 ms	64.292 ms
Average	25.226 ms	40.753 ms
Sample σ	4.89 ms	6.178 ms
95% CI	24.25 ms - 26.2 ms	39.52 ms - 41.97 ms

TABLE III
PATH INSTALLATION TIME USING PROACTIVE FORWARDING.

Description	Until Last Flow_Mod
Minimum	27.125 ms
Maximum	77.752 ms
Average	52.29 ms
Sample σ	13.679 ms
95% CI	49.57 ms - 55 ms

cases did not took advantage of other possible paths available in the network, and load balance the traffic among the links.

For the static path installation measurement, a point to point intent is manually configured on the controller using the ONOS built-in app "push-test-intent" [5]. 100 intents were emulated to measure 100 path installations, where the app execution was made from a remote computer, and its timestamp captured by Wireshark as earlier described in Fig. 2. The time frame between the app execution and the last OFPT_FLOW_MOD message sent by the controller to the switches, is the time response that the controller takes to process the intent submission and install the static path in the network. Table III illustrates the minimum, maximum, and average path installation response time, as well as the sample standard deviation (σ) and the 95% Confidence Interval (CI).

Packet forwarding through the switches had an average delay of 186.49 μs , in which the mean delay is within the range of 178.358 μs and 194.621 μs , with a confidence level of 95%.

B. Segment Routing Testing

The implementation of this method is based on the experimentation made by the Onosproject [6]. In this project, Segment Routing is used as an extension of MPLS into an SDN environment. The same ONOS version utilized for that experimentation (Spring-Open) is also used to conduct the series of tests previously described in the preceding section. The ONOS controller at the network topology is used to load a configuration to the CPqD switches to emulate routing capabilities on them. Each switch is assigned with a loopback IP address and a Segment Routing ID (SID), to identify them as Segment Routing nodes according to [7]. Since Segment Routing is a L3 mechanism, each host is configured under different subnets. Fig. 4 illustrates the logical network topology configured by the controller using the Mininet network topology. In the topology, the flows downloaded to the switches are MPLS and IP table entries that are used by each switch to compute the path between hosts h1 and h2. In all the cases, the initial routes computed were the paths 1 and 2 described in section II, in which both were used during traffic. The forwarding actions are determined by group entries in the group table. Here group chaining is used to push SIDs of the hops involved in the path into the MPLS label stack in order to construct the segment sequence that will route the packet through the network.

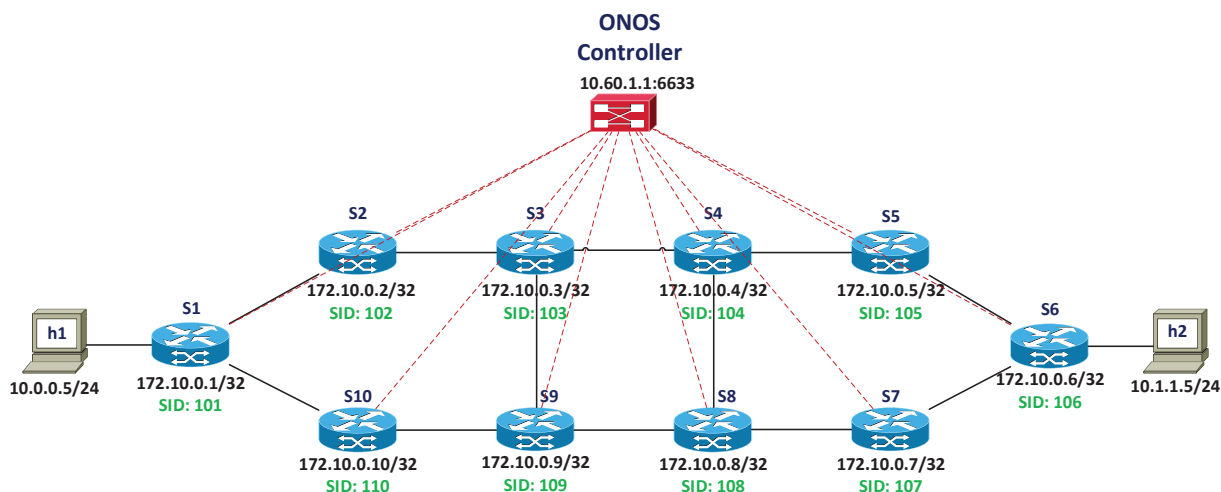


Fig. 4. Segment Routing Topology.

As was done with Proactive Forwarding, the network performance of the Segment Routing topology was tested. The results are presented in Table IV.

Once again, the network topology limits the transmitted bitrate causing huge packet loss over 100 Mbps. The traffic still stable over 1 Mbps, but since the subsequent tests are meant to be compared with the ones of Proactive Forwarding, the bitrate selected for the experimentation is 100 Kbps, the same bitrate for stable traffic in Proactive Forwarding. Using this bitrate the average round trip time observed was 2.734 ms, and an average jitter of 0.0559 ms.

Table V shows the results of the measurement made on the response time in front of network events, in which the emulated failures were made on links s3-s4 and s8-s9 like it was done with Proactive Forwarding, where no matter the path initially calculated, the traffic would pass through the described links. In the initial state of each test, Segment Routing always compute 2 different paths. Thus, it can load balance the traffic using a round robin method within the group action buckets [8].

The time that takes the controller to process the information received by the port status message, and the transmission of the last OFPT_FLOW_MOD message, is the overall time response in front of network events. During the failure recovery, the traffic presented a minimum interruption of an average packet

loss of 0.0034%, with a 95% confidence that the mean jitter is maintained within the range between 0.0526 ms and 0.0613 ms. The processing time is divided in 2 stages, group recovery and path computation. Group recovery is handled by a module of the Segment Routing application called Group Recovery Handler as described in [9]. The buckets within the groups affected by the link failure, are instructed to be deleted from the group by using OFPT_GROUP_MOD messages. The time frame between the port status message and the first OFPT_GROUP_MOD message, is the processing time for the group recovery as illustrated in Fig. 5. It is not clear the instant at which the path computation starts, but the path computation time is some where within the time frame between the first OFPT_GROUP_MOD message and the first OFPT_FLOW_MOD message. For this paper, the overall processing time is assumed to be the time frame between the port status message and the first OFPT_FLOW_MOD message. Table V presents the processing and response time of Segment Routing in front of network events.

In Segment Routing, the path installation time is configured through the implementation of tunnels and policies, that defines a certain path across the network by introducing the related nodes Segment IDs to the MPLS label stack [10]. In this manner, on each hop the destination is set according to the label found in the top of the stack, and then pushed out

TABLE IV
NETWORK PERFORMANCE USING SEGMENT ROUTING.

Bitrate	Packet Loss	Received Bitrate
100 Kbps	0%	100 Kbps
1 Mbps	0%	1 Mbps
10 Mbps	1.6%	10 Mbps
100 Mbps	90%	9.9 Mbps
1 Gbps	98%	9.24 Mbps

TABLE V
RESPONSE TIME USING SEGMENT ROUTING.

Description	Until First Flow_Mod	Until Last Flow_Mod
Minimum	112.232 ms	197.89 ms
Maximum	264.361 ms	435.843 ms
Average	122.399 ms	265.326 ms
Sample σ	20.814 ms	56.432 ms
95% CI	118.27 ms - 126.53 ms	254.13 ms - 276.52 ms

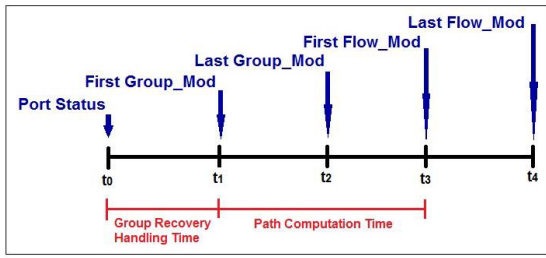


Fig. 5. Segment Routing link recovery process time frame.

at each hop until it reaches the edge node that is connected to the destination network. In this implementation of Segment Routing brought by the ONOS project, the configuration of tunnels represent the arrange of SIDs, which would be the group numbers that would be joined into a group chain. At this point, when the tunnel configuration is executed from the console, the controller processes the creation of the necessary groups and add them to the switches group tables by using OFPT_GROUP_MOD messages. On the other hand, the configuration of policies represents the information of source and destination addresses, the destination group to execute the policy, and the priority that the path would have. When the policies are executed from the console, the controller processes the information into Access Lists (ACL) that are downloaded to the switches by using OFPT_FLOW_MOD messages. Fig. 6 displays the time frame of these processes.

Since the tunnel and policy configurations are separate executions, the measurement cannot be assumed as the complete time frame from the first execution to the last OFPT_FLOW_MOD message, because it would introduce additional delay that is not relevant to the processing time of the controller. In other words, the measurement of the overall time response for the static path installation would be:

$$RT = TCT + PCT, \tag{1}$$

where **RT** is the response time of the path installation, **TCT** is the tunnel configuration time, and **PCT** the policy configuration time.

Table VI illustrates the times measured for the manually installed paths. The measurements presents a sample standard

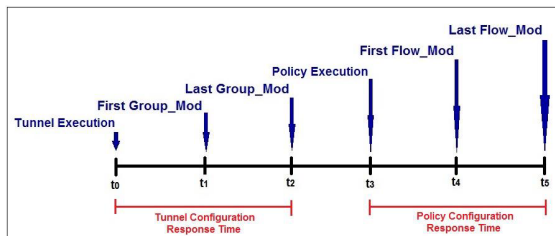


Fig. 6. Segment Routing static path installation time frame.

TABLE VI
PATH INSTALLATION TIME USING SEGMENT ROUTING.

Description	Tunnel	Policy	Path Installation
Minimum	3.42 ms	1.951 ms	5.371 ms
Maximum	9.625 ms	7.104 ms	16.729 ms
Average	7.761 ms	3.69 ms	11.452 ms

deviation (σ) of 2.488 ms, with a 95% confidence that the mean response time is maintained within the range between 10.959 ms and 11.946 ms. Compared with the network event test, the response time is lower, due to the operation that takes place in the process. Thus, it only comprises an installation of an Access List (ACL) entry in the Access List table (ACL table) and the insertion of additional groups in the group table, without modifying the entire group table on the related switches.

The average packet forwarding delay observed on the switch was 290.2 μs and a sample standard deviation of 37.311 μs , with 95% confidence that the mean delay is within the range between 282.796 μs and 297.603 μs . Compared to the test made on Proactive Forwarding, Segment Routing introduces more delay in the switch for packet forwarding.

IV. RESULTS ANALYSIS

The measurements performed are reviewed and analysed on Fig. 7 comparing the results of both mechanism working on the network topology previously described.

There is a clear difference in time response, in which Segment Routing presents more delay in most of the measurements. The switch forwarding delay shows the effects of OpenFlow Pipeline usage introduced by both mechanisms.

On one hand, in Proactive Forwarding, the pipeline processing used is the standard method of OpenFlow as described in [8]. In this procedure, each ingress packet start a lookup for an entry in the first flow table (table 0). It can go to a next flow table or forwarded to an egress port, or to the controller, depending on the action set found in the matched flow entry. In the case of our topology, only one flow table was created

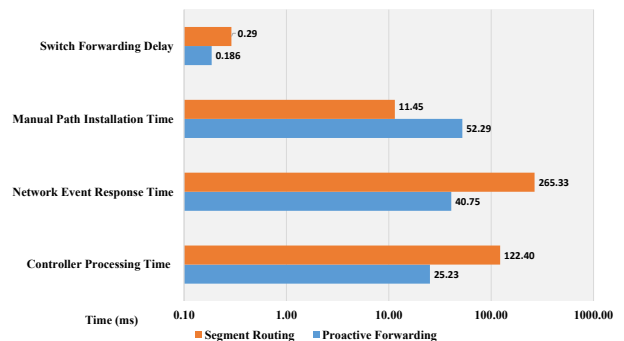


Fig. 7. Test Results Comparison.

with several flow entries. This means that the ingress packet only had to be processed by one table before being forwarded.

On the other hand, Segment Routing uses a series of flow tables plus the group table as described in [9]. In this case, at least 4 tables are used: the IP address routing table, the MPLS table, the Access List table (ACL table) and the group table, which is used to apply the action sets using action buckets within each group. This table is necessary for the pushing of SIDs into the MPLS label stack using the function of group chaining. Fig. 8 shows a summary of the OpenFlow pipeline usage by both path establishment methods.

Segment Routing takes lower time to statically install paths due to the OpenFlow messages transmitted from the controller to the switches. While in Proactive Forwarding the controller sends OFPT_FLOW_MOD and pairs of OFPT_Barrier_[REQUEST—REPLY] messages to install the necessary flows into all related switches, Segment Routing has to send only a few OFPT_GROUP_MOD and OFPT_FLOW_MOD messages into specific switches to establish a path across the network. In most cases, Proactive Forwarding exchange a total of 54 OpenFlow messages, while in Segment Routing exchange only 7 OpenFlow messages. In case of Segment Routing static paths, the OFPT_FLOW_MOD messages introduces new entries to the ACL table, which overrides the instruction set of both MPLS and IP table entries related to specific traffic according to [9].

The opposite situation is observed during the response time in front of network events. Segment Routing has to use both OFPT_GROUP_MOD and OFPT_FLOW_MOD messages, in which the number of messages are greater due to their use in modifying both MPLS and IP table entries (742 OpenFlow messages, including port status and barrier messages). Meanwhile, in Proactive Forwarding only use a set of OFPT_FLOW_MOD messages (206 OpenFlow messages, including port status and barrier messages).

Nevertheless, Segment Routing takes advantage of the available paths in the network. During the experimentation, Segment Routing always used 2 different paths to load balance the traffic across the network. The action buckets within the related groups were processing the packets to forward them to

the egress ports related to each path by using a round robin method. On the other hand, Proactive Forwarding not always was able to compute more than one path to load balance the traffic. In many cases, Proactive Forwarding only computed one path, leaving the rest of the available paths unused.

In overall results, compared with the performance of Proactive Forwarding, Segment Routing introduce an additional response time of approximately 225 ms in front of network events, and 104 μs in terms of switch packet forwarding. While in static path installation, Segment Routing improve its response time in approximately 41 ms. Following the use case of Segment Routing described in this article, this study can be used to monitor the performance of the controller. Moreover, it can be suitable as a starting point to determine what to improve in Segment Routing to obtain a more efficient controller response time. For example, according to the observations, it is known that most of the factors that introduces more response time is the number of OpenFlow messages exchanged between the controller and the switches. In this case, we have several OFPT_FLOW_MOD to change or add flow entries of IP, MPLS, and ACL tables, and OFPT_GROUP_MOD messages to edit an entry of the group table, in which several sets of this type of messages carried repeated instructions to a single switch. This is probably an area to start looking for improvements, in a way to minimize the number of Flows and Group messages without affecting the working principle of Segment Routing.

During the experimentations on failure recovery, the jitter observed was slightly higher compared with the initial testing at steady state conditions. Since all observations were very low (in the order of μs), the time difference is not considered significant. In other words, during network events the traffic doesn't suffer from considerable increase in the jitter.

V. CONCLUSIONS

The results obtained from the entire experimentation are not definitive, it is dependent of several factors like the hardware used for the controller and the switches. Nevertheless, in most cases can be expected the same pattern and behaviour observed in this paper. The study is more focused on the conceptual functionality of both path methods and how their working principle can affect the time response of the controller. For instance, it can be expected that in most cases Segment Routing tends to introduce more time response to the controller on scenarios like failure recovery and packet forwarding, while on the scenario of static path configuration the response time is better. Lets also remember that the implementation of Segment Routing used for this study is still experimental, and several improvements are to be expected in future releases of ONOS.

From the point of view of the network, the main limitation lies on the virtual switch. The CPqD switch works on the user space level, which limits the maximum bandwidth at which the traffic can be forwarded throughout the network. Using a Kernel space switch would have been more realistic, but the implementation used for Segment Routing wouldn't have supported it [11].

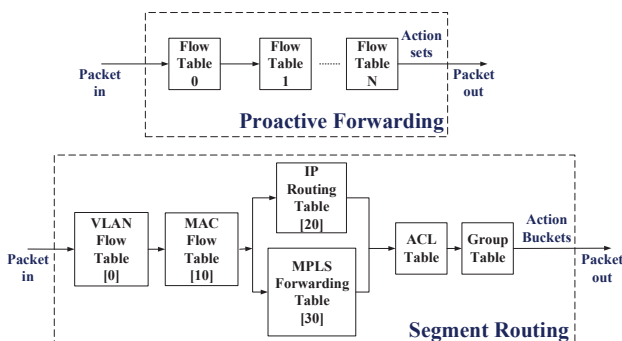


Fig. 8. Proactive Forwarding and Segment Routing O.F. pipeline usage.

Path establishment methods introduce delay to an SDN controller in handling operations related to topology changes, and path configurations, plus an added delay to OpenFlow switches in packet handling. The awareness of these parameters can be used as a guide for performance testing in order to keep track of the time response introduced, and look forward to maintain the desired performance according to the needs established by operators and standards.

The set of tests presented in this paper can be used as a starting point to observe the performance of path establishment methods, their behavior into an SDN environment, and to identify initial needs for improvement that opens the door for further investigation of the tested technologies.

ACKNOWLEDGMENTS

This work has been supported by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2013-47960-C4-1-P.

REFERENCES

- [1] D. C. Frost, B. F. Stewart and C. Filisfil. United States of America Patent US2014/0098675, 2014.
- [2] "OpenFlow," [Online]. Available on May 18th, 2015: <https://www.opennetworking.org/sdn-resources/openflow>.
- [3] P. Berde, M. Gerola, J. Hart, Y. Higuchi, M. Kobayashi, T. Koide, B. Lantz, B. O'Connor, P. Radoslavov, W. Snow and G. Parulkar, "ONOS: Towards an Open, Distributed SDN OS," [Online]. Available on May 18th, 2015: <http://www-cs-students.stanford.edu/~rlantz/papers/onos-hotSDN.pdf>.
- [4] A. Koshibe, "Intent Framework," 2015. [Online]. Available on May 18th, 2015: <https://wiki.onosproject.org/display/ONOS/Intent+Framework>.
- [5] S. Zhang, C. Franke, "Experiment C Plan - Intent Install/Remove/Re-route Latency," 2015. [Online]. Available on Jul 10th, 2015: <https://wiki.onosproject.org/pages/viewpage.action?pageId=3441828>.
- [6] S. Das, "Project Description," 2014. [Online]. Available on May 18th, 2015: <https://wiki.onosproject.org/display/ONOS/Project+Description>.
- [7] S. Das, "Configuring ONOS (spring-open)," [Online]. Available on May 18th, 2015: <https://wiki.onosproject.org/pages/viewpage.action?pageId=2130918>.
- [8] "OpenFlow Switch Specification Version 1.3.4," 2014. [Online]. Available on May 18th, 2015: <https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.3.0.pdf>.
- [9] S. Das, "Software Architecture," 2014. [Online]. Available on May 18th, 2015: <https://wiki.onosproject.org/display/ONOS/Software+Architecture>.
- [10] S. Das, "Using the CLI," 2014. [Online]. Available on May 18th, 2015: <https://wiki.onosproject.org/display/ONOS/Using+the+CLI>.
- [11] S. Das, "Installation Guide," 2014. [Online]. Available on Jul 10th, 2015: <https://wiki.onosproject.org/display/ONOS/Installation+Guide>.

Acceso automático a información museística en dispositivos móviles

Guillermo Azuara¹, Ana María López¹, Ana María Salinas¹, Eduardo Pascual, Jaime Vicente².

¹Departamento de Ingeniería Electrónica y Comunicaciones,
Universidad de Zaragoza.

²Museo provincial de Teruel

EUPT – Ciudad Escolar, s/n. 44.003 Teruel.

{gazuara, lopeztor, salinas}@unizar.es, edu_ftk@hotmail.com, jvicenteredon@dpteruel.es.

Resumen- Este documento presenta el desarrollo de una aplicación para teléfonos móviles con sistema operativo *Android*, que permite obtener información museística de determinadas piezas ubicadas en un museo de manera muy sencilla para el visitante. Se ha utilizado tanto Identificación por Radiofrecuencia (RFID), concretamente comunicación de campo cercano (NFC), como códigos bidimensionales de respuesta rápida (QR codes) para identificar el objeto correspondiente.

Palabras Clave- museos, información museística, teléfono móvil, RFID, NFC y códigos QR.

I. INTRODUCCIÓN

El objetivo perseguido en este trabajo ha sido diseñar y desarrollar un sistema para obtener información de manera automática sobre piezas relevantes del Museo Provincial de Teruel (señaladas por el director, miembro del equipo de investigación) en el dispositivo móvil de un visitante del museo. Con el sistema propuesto, el usuario simplemente debe instalar la aplicación desarrollada en su teléfono móvil (en este prototipo sólo está disponible para el sistema operativo *Android*), y mediante una interfaz muy sencilla e intuitiva con sólo acercar el teléfono o enfocar una etiqueta, se muestra en el dispositivo la información de la pieza (ver Fig. 1).

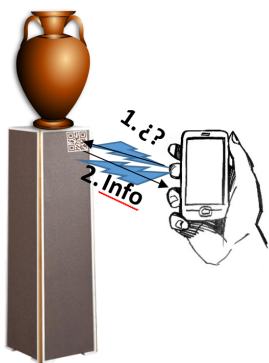


Fig. 1. Esquema del servicio.

Para ello se han utilizado dos tecnologías de Identificación Automática y Captura de Datos (AIDC): los códigos de barras bidimensionales de respuesta rápida (QR Codes) y los chips de comunicación de campo cercano (NFC) que incorporan algunos de los teléfonos móviles.

Los requisitos que se establecieron para la aplicación objetivo del proyecto fueron:

- Funcionar en el dispositivo de los usuarios sin ninguna configuración previa.
- Interfaz intuitivo.
- Funcionamiento automático del sistema: el usuario debe realizar una mínima interacción con su dispositivo para obtener la información, ya que sólo acercándolo o enfocando una imagen el dispositivo deberá deducir de qué pieza se trata y mostrar la información.

En el resto del artículo se presenta un estado del arte, contexto, preparación de la información, aplicación desarrollada y conclusiones y líneas futuras.

II. ESTADO DEL ARTE

Si realizamos en *Google Académico* una consulta sobre la aplicación de tecnologías de Identificación Automática y Captura de Datos (AIDC, Automatic Identification and Data Capture) en los museos, con la cadena de búsqueda “*museum NFC OR RFID OR AIDC*”, podemos comprobar que ha experimentado un creciente aumento en los últimos diez años, como demuestra el número de publicaciones científicas (2000-2005: 1060, 2006-2010: 3840, 2011-2015: 5130).

El término AIDC abarca varias tecnologías utilizadas para la identificación automática de objetos, recoger información sobre ellos e introducir estos datos electrónicamente en sistemas de información [1].

Las tecnologías AIDC existentes incluyen: códigos de barras, identificación por radiofrecuencia, biometría, bandas magnéticas, reconocimiento óptico de caracteres, tarjetas inteligentes y reconocimiento de voz [1], así como diversas variantes y combinaciones de estos.

La principal ventaja de los códigos QR es que la mayoría de los teléfonos móviles inteligentes actuales son capaces de leerlos e interpretarlos, ya que sólo necesitan utilizar la cámara del teléfono y un software que suele estar preinstalado o disponible para su descarga de manera gratuita. Sus principales inconvenientes es que necesitan unas mínimas condiciones de iluminación (dependen mucho de la cámara que integre el teléfono) y que hay que enfocarlos adecuadamente con el dispositivo móvil, por lo que no son aptos para personas con discapacidades visuales.

La primera gran ventaja de los sistemas inalámbricos de identificación es que no hace falta visión directa entre el elemento que contiene la información y el lector; y la segunda que son capaces de almacenar mucha más información y que además ésta se puede modificar de manera dinámica. Su mayor inconveniente es que el dispositivo debe incorporar un lector NFC (Near Field Communication), aunque actualmente ya existen muchos teléfonos en el mercado que los llevan, y se prevé que a muy corto plazo sean introducidos de forma masiva en los nuevos dispositivos.

En la bibliografía existente se han localizado un buen número de publicaciones basadas en el uso de identificación por Radio Frecuencia [2-21]. La mayoría de estas experiencias se basan en el uso de dispositivos específicos, propiedad del museo, por lo que nuestra solución presenta en este sentido una ventaja evidente, ya que al usar dispositivos propiedad del usuario la institución no debe hacer frente ningún coste de mantenimiento ni de adquisición.

En estos trabajos, cuando mencionan la posibilidad de uso de equipos propiedad del usuario [16] se hace desde aplicaciones específicas, que no son sencillas de instalar para un usuario medio, ya que están basadas en Windows Mobile y Symbian. En nuestro sistema, la aplicación que explota la tecnología NFC se ha desarrollado sobre el sistema operativo Android, que según Netmarketshare [22] en Junio de 2015 tenía estimada una cuota de mercado a nivel mundial del 51.04 %, frente al 40.83 % de iOS (iPhone) o el 2.29 % de Windows Phone. Además, la instalación de la aplicación es sumamente sencilla para los usuarios.

La principal ventaja del uso de tecnologías QR es que está disponible en prácticamente todos los teléfonos de gama media y baja (con las limitaciones que comportan el uso de sensores fotográficos de baja calidad), por lo que el sistema es utilizable por la mayoría de teléfonos actuales. Respecto a los trabajos basados en el uso de códigos de barras bidimensionales, como QR, podemos citar [23-25], pero en este caso, dado que se utiliza esta tecnología como único sistema, aparecen problemas inherentes a sus características. Entre ellos podemos destacar que es necesario enfocar con bastante precisión una imagen, y que esta tarea se complica en el caso de lentes sucias, dispositivos con cámaras de poca resolución, etc. Además, la información que pueden almacenar estos códigos es muy reducida, al contrario que otras tecnologías como las etiquetas NFC, que permiten recuperar la información de cada obra de manera local e inmediata, aunque no exista conexión a la base de datos del sistema [26]. Otro ejemplo de utilización de códigos impresos, es su uso en yacimientos arqueológicos, para la identificación de las piezas localizadas [27], aunque en este caso también parte de los miembros del equipo de investigación han implementado con éxito sistemas de identificación por radiofrecuencia para este entorno [28], con todas las ventajas que implica el uso de la tecnología NFC frente a los códigos de barras.

Las características de los sistemas de localización y almacenamiento de información basados en ondas de radio los hacen apropiados para su aplicación en procesos de inventario y logística en diferentes ámbitos sociales y empresariales. Si nos centramos en el mundo de la cultura, es una aplicación muy extendida la instalación de etiquetas en los libros de una biblioteca [29]. Estos sistemas han

permitido agilizar el proceso de localización y seguimiento de cada uno de los volúmenes. Incluso se puede asociar cada uno de ellos a una determinada ubicación que puede ser revelada como errónea simplemente por la lectura de las etiquetas asociadas a dicho lugar y al libro allí situado. Igualmente, los fondos de museos pueden ser gestionados utilizando esta tecnología. La asociación de una etiqueta RFID a cada una de las piezas de arte, el mantenimiento de la base de datos que relaciona esta etiqueta con la obra en cuestión y el diseño de una red de antenas de radiofrecuencia en posiciones estratégicas del edificio del museo, permite conocer de una manera ágil y fiable la localización de cada obra de arte en todo momento [30]. Otras experiencias interesante en un museo con tecnología RFID la encontramos en [31-33].

En [31] a los visitantes se les entrega una etiqueta (en forma de pulsera) RFID y pueden interactuar con algunas de las obras del museo. En [32] se utiliza esta tecnología para controlar el proceso de aprendizaje de los estudiantes en su visita al museo. Gracias al uso de RFID, en [33] se presenta un sistema de recomendaciones sobre qué ver, basado en las visitas previas de otros usuarios del museo.

III. CONTEXTO

Actualmente, uno de los principales problemas de muchas instituciones son los grandes recortes en los presupuestos, y cómo hacer frente a los gastos de mantenimiento de sus instalaciones, donde muchas veces no cabe recorte posible. En este contexto se ha desarrollado un sistema de información museística basado en los dispositivos propiedad de los usuarios (utilizable por la gran mayoría de teléfonos actuales), en este caso los visitantes del museo. Con este enfoque se reducen a cero tanto los gastos de mantenimiento de los dispositivos, como la inversión inicial en adquisición de terminales electrónicos de información.

También se ha incorporado al sistema elementos accesibles, para que personas discapacitadas (ciegas o con movilidad reducida) puedan tener acceso a la información de manera sencilla, simplemente acercando el teléfono a la pieza, podrán ver o escuchar la información asociada. Por los propios requerimientos de hardware esta característica inicialmente sólo estará disponible para teléfonos que incorporen lectores NFC.

Un sistema RFID, NFC es una implementación concreta de esta tecnología, se basa en la posibilidad de intercambiar información entre un lector/escritor RFID y una etiqueta, simplemente cumpliendo los requisitos de proximidad de esa implementación concreta de la tecnología y sin necesidad de visión directa. Esta capa inferior, (ver Fig. 1) , se integra en un contexto determinado, siendo una de las potencialidades del uso de este tipo de tecnología la flexibilidad del *middleware* el sistema.

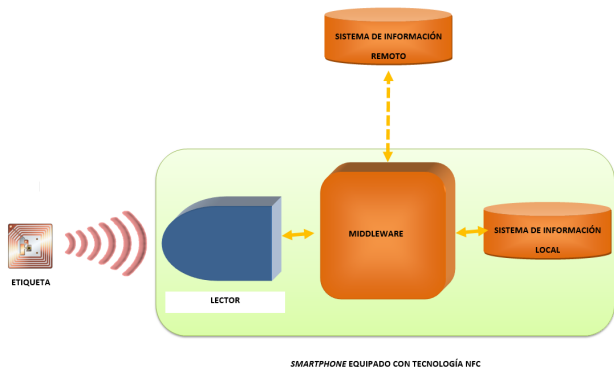


Fig. 2. Ejemplo de funcionamiento.

Como se aprecia en la Fig.2, el uso de la tecnología NFC permite recuperar información directamente de la propia etiqueta, pudiendo además completar esa información con otros recursos, tanto de texto o hipertexto como multimedia, mediante al acceso a una base de datos local o remota vía una red inalámbrica.

El Museo de Teruel es una institución cultural, dependiente de la Diputación Provincial de Teruel, creada con el objetivo de investigar, conservar y difundir el patrimonio cultural turolense y contribuir al conocimiento de las tendencias artísticas generales. Desarrolla programas de arqueología, etnografía y arte contemporáneo, a través de investigaciones y excavaciones arqueológicas, exposiciones y publicaciones específicas. Se encuentra ubicado en la Casa de la Comunidad, uno de los edificios civiles, de estilo renacentista, más emblemáticos de la ciudad de Teruel. Fue construido en 1592, bajo la dirección de los maestros Juan de Rigol y Pedro de Heredia, como sede de las instituciones políticas y jurídicas de la Comunidad, formada por la ciudad de Teruel y sus aldeas. El Museo alberga, distribuidas por las seis plantas del edificio, importantes colecciones de arqueología, que ofrecen una completa visión de las formas de vida de las poblaciones antiguas de la provincia de Teruel, desde la Prehistoria hasta la Edad Media, destacando las colecciones de cultura ibérica y romana; y de Etnografía, que permiten al visitante conocer las formas de vida tradicionales en aspectos como la vivienda, el mobiliario, los recursos económicos, las principales actividades económicas, o la indumentaria y las costumbres religiosas. Además cuenta con una importante colección de arte contemporáneo, proveniente en su mayoría de las Becas Endesa, no expuesta en la actualidad. Las colecciones del Museo destacan por su variedad y complementariedad.

Para este prototipo se ha trabajado con un total de 14 piezas, algunas ellas de excepcional valor arqueológico. Aunque el museo cuenta con cartelas en las piezas y explicaciones gráficas, las nuevas tendencias en musealización de colecciones apuestan por dar todo el protagonismo a las piezas, minimizando la cartelería y centrándose en esquemas generales para la vitrina. El sistema propuesto se integra perfectamente con esta filosofía, ya que de una manera muy limpia (con una pequeña etiqueta NFC que también lleva impreso un código QR) se puede recuperar y ampliar la información de las piezas más importantes.

Finalmente señalar, que el sistema desarrollado es fácilmente escalable y exportable a otros entornos museísticos.

IV. PREPARACIÓN DE LA INFORMACIÓN

Una parte importante del trabajo ha sido analizar la información que disponía el museo de cada una de las piezas y rediseñar la base de datos para almacenar dicha información.

En la base de datos de la aplicación cada registro asociado a una pieza contiene los siguientes campos: N° de pieza, N° de inventario, Objeto, Clase, Contexto, Datación, Localización, Observaciones, URL, Imagen.

También se realizó un importante esfuerzo para minimizar la cantidad de memoria utilizada por cada registro, para lo que se crearon tablas de apoyo para poder codificar los valores más habituales de algunos de los campos, logrando un importante ahorro (25 %) de memoria en la base de datos que se instala con la aplicación. Se decidió utilizar esta metodología para codificar los campos: clase, contexto y localización.

Las tablas utilizadas han sido:

- a) principal: tabla principal que contiene la información básica de un objeto (ver Fig. 3).
- b) clases: enumera todas las clases genéricas dadas de alta en el sistema (Ej.: Objetos bélicos y militares, Ganadería y caballería, Útiles de mezcla o batido, Equipamiento doméstico,...). Cada una de estas clases se identifica con un número que, posteriormente, se utilizará para asociar un objeto con su clase genérica. Puede darse el caso de que un objeto tenga dos clases genéricas asociadas, pero como máximo, solamente dos.
- c) contextos: enumera todos los contextos culturales dados de alta.
- d) localizaciones: enumera los yacimientos de la provincia dados de alta en el sistema.
- e) clase_objeto: relación de como máximo dos clases genéricas con un objeto individual.
- f) contexto_objeto: relación de como máximo dos contextos culturales con un objeto individual.

Tabla principal

_id	uid	inventario	objeto	clase	contexto	datacion	localizacion	observaciones	url	imagen
Auto_increment integer	text	text	text	integer	integer	text	integer	text	text	integer

↑
Clases

↑
Contextos

↑
Localizaciones

Fig. 3. Tabla principal de la base de datos.

V. APLICACIÓN DESARROLLADA

La aplicación, gratuita, ha sido desarrollada para el sistema operativo Android, y básicamente permite intercambiar información entre una etiqueta, ya sea impresa (QR) o con un chip integrado (NFC), y el terminal móvil del usuario. También se incluye información básica del museo como horarios, historia, localización, etc.

Cada etiqueta, asociada con una pieza del museo, contiene una información, un número de inventario, que

permite a la aplicación acceder a una base de datos local desde donde recuperar toda la información complementaria que mostrar al usuario.

Desde el menú principal de la aplicación (Fig. 4) se puede acceder a información general del museo, el plano, el menú de configuración y ayuda de la aplicación. Pulsando los botones “lectura NFC” o “lectura QR” el usuario elige el método de captura de datos que va a utilizar (lógicamente si el hardware de su teléfono lo permite).



Fig. 4. Menú principal de la aplicación.

El diagrama de la aplicación se muestra a continuación (Fig. 5):

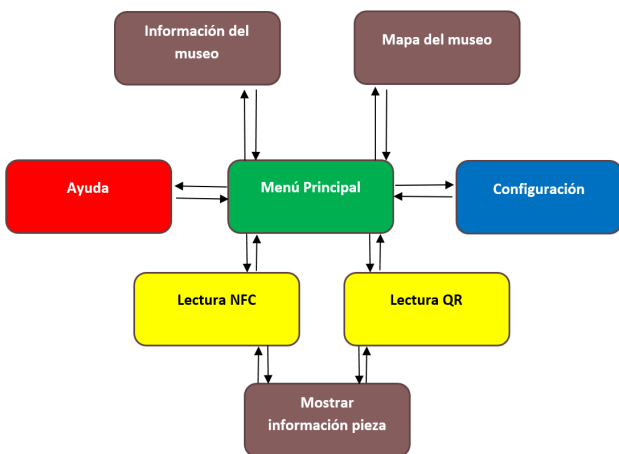


Fig. 5. Diagrama de la aplicación.

El motor de base de datos utilizado es SQLite. Se trata de un motor muy popular en la actualidad por ofrecer características tan interesantes como su pequeño tamaño, no necesitar servidor externo dedicado, precisar poca configuración, ser transaccional y, por supuesto, ser de código libre. Para la creación de la base se ha utilizado el complemento *SQLite Manager*, ofrecido por *Mozilla* para el navegador web libre *Firefox*.

La base de datos está compuesta por las seis tablas enumeradas en el apartado IV que contienen toda la información relacionada con los objetos dados de alta en el sistema.

En la siguiente captura (Fig. 6) se muestra un ejemplo de la información que se visualiza tras haber identificado una pieza (la interfaz incorpora los modos habituales de desplazamiento con los gestos convencionales en este entorno).



Fig. 6. Ejemplo de información visualizada en la aplicación.

Para garantizar la coherencia de la información asociada a las diferentes piezas, la base de datos contiene un trigger denominado *localizacion_princ* que comprueba que cuando se inserta una nueva localización en la tabla principal está dada de alta con anterioridad en la tabla localizaciones. También verifica la integridad referencial de la base de datos.

Con la versión actual y dado que no se cuenta con conexión a un servidor remoto, las imágenes que se muestran en la aplicación son accedidas en modo local, es decir se almacenan en el terminal en el proceso de instalación de la aplicación.

Actualmente, y debido a la imposibilidad de utilizar infraestructura externa de interconexión de redes, la base de datos es estática, es decir, que una vez cargada en la estructura de la aplicación no puede actualizarse dinámicamente.

El proceso de actualización de la base de datos consistiría en hacer los cambios a través del gestor de base de datos mencionado anteriormente y luego descargar la nueva versión de la aplicación (de manera similar a cómo el usuario actualiza cualquier aplicación de *Google Play*).

VI. CONCLUSIONES Y LÍNEAS FUTURAS

El funcionamiento del sistema se basa en dos maneras de obtener la identificación de las piezas de las que se requiere información: mediante códigos QR y mediante el uso de tecnologías inalámbricas de identificación de objetos. Al implantar simultáneamente ambas tecnologías se obtienen las ventajas de las dos, salvando así mismo sus inconvenientes.

El resultado principal del proyecto ha sido el prototipo del sistema (constituido por la aplicación y la infraestructura asociada).

Una fase muy importante del proyecto ha sido el análisis de las bases de datos existentes, para minimizar la memoria utilizada en el dispositivo móvil y mejorar la estructuración de la información. Con el uso de códigos numéricos en vez de cadenas alfanuméricas se ha conseguido un ahorro de memoria en torno al 25 %.

En la versión del prototipo que finalmente se ha implementado y dado que no fue posible el despliegue de una red local inalámbrica tal como se había planteado inicialmente, se ha realizado la aplicación con los datos autocontenidos, de manera que no necesita conexión a ningún servidor y por tanto tampoco utilizar ningún tipo de conexión de datos.

Una de las partes más importantes, fue la de las pruebas in-situ de las etiquetas. Hubo que realizar diferentes mediciones y pruebas para obtener resultados satisfactorios con ambas tecnologías, ya que por los materiales de construcción de los expositores, como por las condiciones de iluminación, se presentaron problemas tanto con las etiquetas NFC como con los códigos de barras bidimensionales. Se realizaron diversas pruebas de funcionamiento:

- Códigos QR: diferentes condiciones de iluminación, ubicación, colores y tamaños.
- Etiquetas NFC: diferentes condiciones de ubicación y medición de interferencias con materiales (especialmente metales).
- Diferentes usuarios y diversos dispositivos móviles.

Tras la evaluación del prototipo, los beneficios que puede brindar la implantación del sistema son:

- Revalorización de la visita al museo.
- Mejora de la imagen del museo, gracias a las nuevas tecnologías y a la accesibilidad para personas discapacitadas. Mejora de la calidad de las visitas.
 - Posibilidad de exportación a otros entornos.
 - Facilidad en el traslado de piezas y colecciones (tanto dentro del museo, como en su cesión a otros museos): el sistema es fácilmente reubicable.
 - Sistema fácilmente ampliable.
 - La información viaja con la propia pieza sin necesidad de elementos externos explicativos.
 - Localización y control de las piezas.
 - El uso de las dos tecnologías citadas hace que el sistema pueda ser utilizado por la mayoría de teléfonos actuales.

Como conclusión señalar que los diversos tipos de usuario que han probado el prototipo para la realización de las pruebas (investigadores, personal del museo, periodistas,...) mostraron su satisfacción con la aplicación, y

sobre todo con las posibilidades que abre de cara a mejorar las visitas al museo.

El siguiente paso lógico sería distribuir la aplicación, que ya es totalmente funcional, y recopilar información de su uso, tanto cuantitativa como cualitativa.

Las mejoras de la aplicación pasan necesariamente por el despliegue de una infraestructura de red inalámbrica que permita la conexión con un servidor remoto, que permita tanto una actualización más sencilla por parte de los responsables del museo, como un menor tamaño de la aplicación y la posibilidad de utilizar otro tipo de información multimedia (vídeos, infografías,...) e interacción con el usuarios (opiniones, sugerencias, etc...).

En la línea del punto anterior, se desarrollará un interfaz web para añadir y modificar registros en la base de datos centralizada, desde los equipos de trabajo del personal del museo. Se desarrollará un manual de usuario para los usuarios con permisos de gestión de la base de datos.

Con la base de datos centralizada, se puede aumentar la información relativa a las piezas (más fotografías, detalles, imágenes de la excavación, infografías interactivas, vídeos ...).

También se considera conveniente la exportación de la aplicación a dispositivos iOS.

AGRADECIMIENTOS

Fundación Universitaria Antonio Gargallo (proyecto 2012/C002) y Caja de Ahorros de la Inmaculada.

REFERENCIAS

- [1] A.D. Smith and F. Offodile, "Information management of automatic data capture: an overview of technical developments", in *Information Management & Computer Security*, 10(3), 2002, 109-118.
- [2] F. Cafaro, L. Lyons, J. Radinsky and J. Roberts, "RFID localization for tangible and embodied multi-user interaction with museum exhibits," in *12th International Conference on Ubiquitous Computing, UbiComp 2010, Copenhagen, 2010*, pp. 397-398.
- [3] D. Cai, "Development of an interactive museum guidance system with ubiquitous communication," in *10th IASTED International Conference on Software Engineering, SE 2011, Innsbruck, 2011*, pp. 111-115.
- [4] T. - Chen, V. Istanda and C. -. Kuo, "The case study of RFID innovative application services in national palace museum," *ICIC Express Lett.*, vol. 5, pp. 207-212, 2011.
- [5] A. Fevgas, P. Tsompanopoulou and P. Bozanis, "iMuse mobile tour: A personalized multimedia museum guide opens to groups," in *16th IEEE Symposium on Computers and Communications, ISCC'11, Corfu, 2011*, pp. 971-975.
- [6] A. Fevgas, P. Tsompanopoulou, A. Tsiouvolos, G. Drasidis and P. Bozanis, "Utilizing UHF RFIDs to enhance museum visiting experience," in *14th Panhellenic Conference on Informatics, PCI 2010, Tripoli, 2010*, pp. 16-20.
- [7] H. -. Hsu and H. -. Liao, "A mobile RFID-based tour system with instant microblogging," *J. Comput. Syst. Sci.*, vol. 77, pp. 720-727, 2011.
- [8] Y. -. Huang, Y. -. Chang and F. E. Sandnes, "Experiences with RFID-based interactive learning in museums," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 3, pp. 59-74, 2010.
- [9] Y. -. Huang, S. -. Wang and F. E. Sandnes, "RFID-based guide gives museum visitors more freedom," *IT Prof.*, vol. 13, pp. 25-29, 2011.
- [10] H. -. Kim, C. J. Park and S. B. Kim, "Improvement of jeju national museum service using mobile RFID," in *2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ICIS 2009, Seoul, 2009*, pp. 1066-1071.
- [11] C. -. Kuo and J. K. Chiang, "The preliminary study of ubiquitous infrastructure of museum service applications in national palace museum," in *9th WSEAS International Conference on Applied Computer and Applied Computational Science, ACACOS '10, Hangzhou, 2010*, pp. 50-54.

- [12] C. -. Kuo and J. K. Chiang, "A ubiquitous infrastructure applications to support museum's service," in IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2009, Hong Kong, 2009, pp. 2423-2425.
- [13] C. -. Kuo, J. K. Chiang and Q. -. Lin, "The research on applying RFID information system architecture for museum service," in 4th International Conference on Computer Sciences and Convergence Information Technology, ICCIT 2009, Seoul, 2009, pp. 218-222.
- [14] C. -. Kuo and S. -. Chou, "The ubiquitous museum service environment: Concept, design, implementation, and a case study," in 5th International Conference on New Trends in Information Science and Service Science, NISS 2011, Macau, 2011, pp. 233-236.
- [15] C. -. Kuo and S. -. Chou, "The research of ubiquitous infrastructure of museum service applications in National Palace Museum," WSEAS Trans. Inf. Sci. Appl., vol. 7, pp. 880-889, 2010.
- [16] A. Kuusik, S. Roche and F. Weis, "SMARTMUSEUM: Cultural content recommendation system for mobile users," in 4th International Conference on Computer Sciences and Convergence Information Technology, ICCIT 2009, Seoul, 2009, pp. 477-482.
- [17] N. Li, W. Ni, H. Cheng and J. Tan, "A systematic RFID application platform with integration capability for tour and exhibition," in 2011 2nd IEEE RFID Technologies and Applications Conference, RFID-TA 2011, Collocated with the 2011 IEEE MTT-S International Microwave Workshop Series on Millimeter Wave Integration Technologies, IMWS 2011, Sitges, 2011, pp. 37-41.
- [18] A. Mody, M. Akram, K. Rony, S. A. Muhammad and R. Kamoua, "Enhancing user experience at museums using smart phones with RFID," in 2009 IEEE Long Island Systems, Applications and Technology Conference, LISAT 2009, Farmingdale, NY, 2009, .
- [19] I. Satoh, "Design and implementation of context-aware museum guide agents," IEICE Trans. Inf. Syst., vol. E93-D, pp. 789-799, 2010.
- [20] J. Shang, S. Yu, F. Gu, Z. Xu and L. Zhu, "A mobile guide system framework for museums based on local location-aware approach," in 2011 International Conference on Computer Science and Service System, CSSS 2011, Nanjing, 2011, pp. 1935-1940.
- [21] Tsung-Yu Liu, Tan-Hsu Tan and Yu-Ling Chu, "The ubiquitous museum learning environment: Concept, design, implementation, and a case study," in Advanced Learning Technologies, 2006. Sixth International Conference on, 2006, pp. 989-991.
- [22] Netmarketshare, <http://www.netmarketshare.com>. Última consulta Julio de 2015. [on-line]
- [23] U. B. Ceipidor, C. M. Medaglia, A. Perrone, M. D. Marsico and G. D. Romano, "A museum mobile game for children using QR-codes," in 8th International Conference on Interaction Design and Children, IDC 2009, Como, 2009, pp. 282-283.
- [24] Y. -. Huang, Y. -. Chang and F. E. Sandnes, "QR code data type encoding for ubiquitous information transfer across different platforms," in Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing in Conjunction with the UIC'09 and ATC'09 Conferences, UIC-ATC 2009, Brisbane, 2009, pp. 292-297.
- [25] S. Levaldi Ghiron, C. M. Medaglia and A. Perrone, "'art-sonomy': Social bookmarking of real artworks via mobile applications with visual tags," Lect. Notes Comput. Sci., vol. 5616 LNCS, pp. 375-384, 19 July 2009 through 24 July 2009, 2009.
- [26] J. Banks, M. A. Pachano, L. G. Thompson and D. Hanny, RFID Applied. Hoboken, New Jersey: Wiley, 2007.
- [27] J. Martínez-Moreno, P. González Marcén and R. Mora Torcal, "Data matrix (DM) codes: A technological process for the management of the archaeological record," Journal of Cultural Heritage, vol. 12, pp. 134-139, 0, 2011.
- [28] Ana María López, Ana María Salinas, Eduardo Pascual, Guillermo Ignacio Azuara, Gloria Fernández, Elena Gallego, Francisco Burillo, "Identifying and Tracing Archaeological Material with RFID Tags", in Archaeology in the Digital Era, pp 55-64, Amsterdam University Press, 2014. [26] S. Hsi and H. Fait, "RFID enhances visitors' museum experience at the Exploratorium", Commun ACM, vol. 48, pp. 60-65, 2005.
- [29] D. Molnar and D. Wagner, "Privacy and security in library RFID issues, practices, and architectures", in Proceedings of the 11th ACM Conference on Computer and Communications Security, CCS 2004, Washington, DC, pp. 210-219, 2004.
- [30] "The Art of Tracking Masterpieces", RFID Journal, vol. Junio 2002, 2002.
- [31] S. Hsi and H. Fait, "RFID enhances visitors' museum experience at the Exploratorium" Communications of the ACM, vol. 48 (9), pp. 60-65, 2005.
- [32] Chia-Chen Chen and Tien-Chi Huang, "Learning in a u-Museum: Developing a context-aware ubiquitous learning environment", in Computers & Education, Vol. 59 (3), pp. 873-883, 2012.
- [33] Y.D. Seo and J.H. Ahn, "Efficient NFC Tagging Pattern-Based Contents Recommendation for Museum Viewers", in International conference, Information technology and management innovation, 2013.

Modelo para la ubicación de servicios en Cloud Híbrido

Adrià Mallorquí, Agustín Zaballos, Joan Navarro, Ramon Martín de Pozuelo, Alan Briones

Departamento de Ingeniería
La Salle - Universidad Ramon Llull (URL)
08022 Barcelona

{amallorqui, zaballos, jnavarro, ramonmdp, abriones}@salleurl.edu

Resumen- El uso de *clouds* híbridos permite que las empresas puedan cubrir su demanda de recursos IT ahorrando en costes y ganar flexibilidad en el despliegue de infraestructuras, mediante un pago bajo demanda de dichos recursos. Sin embargo, delante de un escenario con distintos servicios a ubicar en más de un *cloud*, es necesario encontrar la mejor distribución de servicios posible para tener el menor coste. Con la revisión de otros estudios se ha podido observar que hay una tendencia a encontrar la mejor distribución de los servicios basándose en el uso de una serie de métricas para el cálculo de este coste. Este artículo realiza una síntesis de diferentes métricas previamente estudiadas y propone una estrategia para definir un modelo de coste para evaluar las ubicaciones de servicios en diferentes *clouds* disponibles. Para ello, y teniendo en cuenta la cantidad ingente de posibilidades, se parte de una prueba de concepto en un escenario con un número limitado de combinaciones. Dado dicho escenario simple de referencia, se presenta un algoritmo que calcula la mejor distribución de servicios posible de manera automatizada y es comparado con otras soluciones existentes.

Palabras Clave—Cloud híbrido, Ubicación de recursos, Modelo de coste

I. INTRODUCCIÓN

La computación en la nube (*cloud computing*) es un paradigma que ofrece, bajo demanda, servicios de computación en Internet a los clientes. Utilizando granjas de servidores, se ofrecen servicios -de computación y también de almacenamiento (*cloud storage*)-, permitiendo un acceso remoto del cliente a esta información o aplicación. Así pues, los servidores son capaces de atender las peticiones en cualquier momento, pudiendo acceder los clientes a dicha información desde cualquier dispositivo y lugar, con conexión a Internet, ofreciendo una gran movilidad y escalabilidad según el modelo de despliegue.

Básicamente existen tres tipos de modelos de despliegue del *cloud*. El despliegue de un *cloud* público se caracteriza pues en ofrecer los recursos requeridos por cualquier empresa sobre infraestructuras compartidas beneficiándose de las economías que se aplican al amortizar la infraestructura global con múltiples clientes. A estos recursos se accede a través de conexiones de datos creando plataformas elásticas que pueden crecer y decrecer dependiendo de las necesidades de cada momento, pagando únicamente por los recursos utilizados, pudiendo absorber picos de demanda sin necesidad de sobredimensionar los recursos. La ventaja principal es pues su funcionamiento bajo demanda. Los *clouds* públicos comerciales son plataformas altamente extendidas en la actualidad, con compañías y plataformas referentes como Amazon o Microsoft [1], que

ponen a disposición del cliente sus servidores para poder desplegar allí sus servicios.

También existe el modelo de *cloud* privado, donde la plataforma *cloud* construida sólo es explotable para la propia empresa que la ha desplegado. Dicho de otra forma, el *cloud* privado es el que una institución tiene acceso exclusivo a los recursos que se han utilizado para implementarlo con funciones análogas a los *datacenter* internos de que disponen algunas empresas, con infraestructura y máquinas propias, dimensionadas en base a la demanda esperada [2]. Estas soluciones generan una sensación de mayor seguridad para los usuarios que disponen de este tipo de despliegues, motivado por el hecho de no compartir recursos con otros usuarios y al hecho que normalmente se despliega considerando tecnologías afines a las propias de la empresa. Sin embargo, el despliegue puede estar fuera del alcance de muchas empresas, pues a pesar de su mayor capacidad de personalización, requiere de fuertes inversiones tanto iniciales como en su operación y actualización que el cliente debe asumir para evitar su obsolescencia. Este hecho no sucede con el *cloud* público, donde dichos costes los asume el proveedor.

Durante los últimos años el uso de la computación *cloud* por parte de las empresas ha ido creciendo y cada vez son más las organizaciones que deciden adoptar una arquitectura de *cloud* híbrido, buscando las ventajas de *clouds* privados y públicos [3]. Este modelo combina recursos de ambos modelos aprovechando la propia infraestructura que disponen algunos clientes, pero insuficiente para poder responder satisfactoriamente ante picos de trabajo, que pueden ser abastecidos por un proveedor externo. El ahorro y la elasticidad que ofrecen los servicios de *cloud* público son mayores que los de *clouds* privados. Por lo tanto, una organización puede optimizar al máximo su eficiencia utilizando servicios *cloud* públicos para las operaciones que no tengan carácter crítico, y recurrir únicamente al *cloud* privado en los casos en que sea realmente necesario [4]. El problema principal de esta opción es su complejidad pues requiere interconectar e integrar una infraestructura propia con otra gestionada por otra institución. Por ello, los *clouds* híbridos generalmente requieren niveles mayores de gestión y administración [5] y requieren de un sistema decisor, orquestador de recursos, para la automatización de su gestión. Además, puesto que los diferentes recursos se ubican en *datacenters* diversos y son gestionados por organizaciones diferentes, es importante asegurar la visibilidad del conjunto a través de alguna herramienta.

Aunque existen muchos casos prácticos para la computación *cloud* híbrida, este estudio tomará como referencia el caso planteado en el proyecto europeo FINESCE [21] en el que se plantea un *cloud* privado para guardar aquellos datos recientes más sensibles generados en un entorno Smart Grid (datos de contadores, puntos de recarga de vehículo eléctrico, etc.), pero que éste no siempre es suficiente para mantener los datos históricos. Se recurre pues a un *cloud* público para responder a los picos de demanda, utilizando un *burst outsourcing*. En el caso del *burst outsourcing*, solamente existe un gasto adicional bajo demanda cuando el *cloud* privado no puede suministrar todos los servicios, momento en que el/los *cloud/s* público/s proveen de los recursos necesarios complementarios [6].

Sin embargo, el factor crucial en el ahorro de costes económicos en IT usando un *cloud* híbrido es la ubicación óptima de recursos. Si imaginamos un escenario con distintos servicios que deben ser ubicados en más de un *cloud*, la elección de la distribución de estos servicios no es trivial y un tiempo de decisión elevado puede ir en detrimento de la QoS ofrecida por el *cloud*. El hecho de elegir una ubicación determinada sin tener una estrategia definida y correcta puede que comporte la elección de una distribución de recursos que no sea la mejor para cumplir con los requisitos que se hayan definido o tenga un coste muy superior al óptimo. Es necesario, por lo tanto, diseñar una estrategia mediante una serie de reglas que marquen las preferencias, prioridades y los límites de costes, temporales, etc. con el fin de obtener la mejor ubicación de servicios posible en un escenario concreto. El primer paso para realizar este proceso es analizar los distintos *clouds* en base a métricas objetivas medibles.

En la Sección II de este artículo se repasa brevemente esta estrategia basada en métricas. La Sección III se explica brevemente la estrategia basada en los modelos de coste económico para presentar la propuesta de un modelo que ayude a obtener una ubicación de servicios óptima dado un caso concreto. En la sección IV se presenta el algoritmo que calcula la mejor ubicación posible de forma automatizada, llamado *Cost Model Based Service Placement Finder* (CMBSPPF). En la sección V se evalúan los resultados obtenidos presentando un escenario de ejemplo para, finalmente, resumir las conclusiones del estudio realizado.

II. ESTRATEGIA BASADA EN MÉTRICAS

Como primera contribución, este artículo presenta una tabla (Fig.1) con los datos obtenidos fruto de su trabajo en el proyecto de investigación [21]. Ésta define de forma cualitativa la importancia de varias métricas en un *cloud* según qué acción se esté llevando a cabo, con una ponderación de 1 (poca importancia), 2 (importancia media) ó 3 (mucha importancia). Destacar también el código de colores utilizado, en el cual las casillas en rojo (1), amarillo (2) o verde (3) fueron definidas mediante un proceso de opinión cruzada entre diferentes grupos de interés implicados en el desarrollo del proyecto, y las casillas en azul fueron posteriormente modificadas en una segunda iteración, analizando las aportaciones y opiniones de otros socios participantes en dichos proyectos (proveedores de *cloud*).

Partiendo de esta tabla, y tras la revisión de varios estudios similares [3][7][8][9], se ha comprobado que existe una tendencia a seguir una estrategia concreta para ubicar los

servicios de la mejor forma posible. Esta estrategia se basa en utilizar métricas para evaluar la ubicación de los servicios en un *cloud* u otro.

Algunos estudios [7][8] parten de la idea que la ubicación de recursos siempre será más barata en el *cloud* privado que en los *clouds* públicos, ya que se supone de inicio que ya se disponen de los recursos del *cloud* privado (la inversión para crear el *cloud* ya ha sido ejecutada con anterioridad). De este modo, la filosofía propuesta es ubicar todos los servicios en el *cloud* privado (del cual no se deberá asumir ningún coste adicional salvo la propia operación) hasta un valor umbral a partir del cual los recursos deberían ser ubicados en el *cloud* público al aparecer picos de trabajo con un coste bajo demanda asociado. Para definir este valor umbral es necesario el uso de métricas que ayuden a marcar el límite de utilización del *cloud* privado. A modo de ejemplo y para ilustrar este concepto, en [7] se toma como referencia el *deadline* del servicio a ubicar y en [8] la métrica usada es el *workload* acumulado.

A. Deadline

El *deadline* es el tiempo máximo para realizar una determinada tarea. En este caso, cuanto más amplio sea el *deadline*, la tarea será menos crítica y requerirá menos recursos. Por esta razón se recomienda ubicar en el *cloud* privado las tareas con un *deadline* menos restrictivo. A medida que el *deadline* va disminuyendo, la demanda de recursos puede ir aumentando y en el momento en que dicha demanda sea excesiva para el *cloud* privado se deberá ubicar dicha tarea en el *cloud* público, donde se podrán obtener los recursos necesarios de forma escalable y tener más garantías de cumplir con el *deadline*. En [7], después de haber realizado varias pruebas, se considera que el valor umbral del *deadline* está alrededor de las 10 horas, por lo que cuando se disponga de menos de 10 horas la ubicación deberá ser pública.

B. Workload

El *workload* es la carga de trabajo que tiene el *cloud*. En estudios previos [7][8] se plantea que cuando una determinada tarea supone incrementar la carga de trabajo de tal forma que el porcentaje de *workload* acumulado supera un cierto umbral (*base workload*), dicha tarea se deberá ubicar en el *cloud* público. Mientras no se supere este umbral, las tareas se deberán ubicar en el *cloud* privado, ya que éste aún las podrá manejar al no suponer un nivel suficiente de estrés como para penalizar otras tareas.

En [8] se define que el límite gestionable por el *cloud* privado es, aproximadamente, el 40% del máximo. Momento en el que aparecen picos de carga (*flash crowd workload*) que deberán ser asumidos por el *cloud* público, el cual podrá aumentar sus recursos bajo demanda. De todos modos el valor umbral del *workload* variará según la capacidad que tenga el *cloud* privado del que se dispone. Aunque la predicción del *workload* máximo no es trivial, en [10] se presenta el algoritmo *FastTopK*, capaz de realizar una estimación fiable.

C. Unión de métricas

Aunque el uso y análisis de una única métrica puede ayudar a ver cómo es el comportamiento de la arquitectura híbrida, el siguiente paso debería ser la combinación de más de una métrica para calcular la ubicación óptima del recurso.

Aunque idealmente el hecho de emplear más de una métrica supone un mejor conocimiento del comportamiento del *cloud*, y por lo tanto una mejor estrategia de ubicación, esto conlleva que la elección del umbral no es del todo trivial ya que se debería crear un árbol de decisiones ponderando el peso de cada una de las métricas utilizadas y teniendo en cuenta la afectación que podrían tener unas con las otras. Este es el motivo por el cual la mayoría de estudios acaban decantándose por utilizar la métrica que se considera más determinante y que da una percepción del umbral. Por lo general la definición de este umbral tiende siempre a ser más subjetiva que la evaluación de datos numéricos, ya que el límite de cada *cloud* privado es variable, e incluso de un mismo *cloud* en distintos escenarios según los intereses del administrador del sistema.

De hecho, aunque existen diversos estudios realizados centrados específicamente en cómo clasificar las métricas de un *cloud* [11][12][13][14], también se han presentado algunas propuestas sobre la manera de computar de forma global el servicio *cloud* [15], con el fin de extraer una única figura de valor que determine el comportamiento adecuado de éste. En cuanto a la metodología de evaluación de los sistemas, existen estudios [12] que más allá de clasificar las métricas, clasifican la propia manera de evaluar los sistemas y extraer estas métricas. Además, pueden aportar matices entre una métrica u otra, o incluso entre dos mediciones de una misma métrica, según la manera y el entorno en que se ha extraído. Se puede realizar una clasificación inicial como la de la Figura 1 que agrupa las métricas relacionadas con el propio *cloud* en cuanto a computación y almacenamiento, y las métricas relacionadas con la red. Con el objetivo de complementar esta clasificación cualitativa, se ha añadido otro conjunto de métricas consideradas como miscelánea, en el que se engloban factores económicos, elasticidad, escalabilidad o seguridad.

D. Síntesis

Así pues, en esta sección se han podido identificar numerosos artículos que abarcan el estudio de las métricas en el *cloud*. La respuesta a cómo identificar, clasificar, medir y valorar la importancia de estas métricas son las aportaciones de estos estudios que han sido tomados como referencia para la confección de la Figura 1. Una vez identificadas las métricas, el siguiente paso consiste en utilizarlas para poder dar soluciones óptimas a otros problemas que en el caso de este estudio es la ubicación de recursos en un *cloud* híbrido. Sin embargo, en la revisión del estado del arte (todos los estudios previamente mencionados) se ha detectado que en la práctica existe una enorme dificultad para poder realizar una toma de decisiones teniendo en cuenta un conjunto de métricas, debido a la dependencia que hay entre ellas. Además, esto también supone una mayor dificultad en la definición de un valor umbral que tenga en cuenta todo el conjunto de métricas y que ayude a definir a partir de qué estado los recursos pasan a ubicarse en un *cloud* u otro. Los estudios analizados hasta la fecha sólo definen valores umbrales para una única métrica (p.ej. el *deadline* y el *workload*).

Después de entender los problemas del uso de un conjunto de métricas para la definición de un valor umbral, en este estudio se observó que en la Figura 1 únicamente las métricas de coste energético, coste de mantenimiento, coste de la propiedad y el coste económico general tenían una

importancia medio o alta en todos los aspectos (en el resto de métricas hay alguna aplicación con importancia baja). Por este motivo, se ha optado por proponer un modelo de coste que englobara todos estos aspectos que, en caso de estar bien definido, podría dar una valoración global y no exclusivamente respecto a un único proceso del *cloud*. Esta solución, aunque directamente no utilice las métricas propias de red o del *cloud*, sí que utiliza métricas que aparecen en la Figura 1 (métricas de coste), que pueden tener una afectación sobre las otras.

III. PROPUESTA DE UN MODELO DE COSTE

La estrategia basada en modelo de coste económico es un método que se basa en la evaluación de datos numéricos objetivos en vez de definir un valor umbral. Esta estrategia consiste en definir un modelo de coste económico preciso que permita calcular numéricamente cual será el coste económico de ubicar los servicios de una forma determinada frente a un escenario planteado en donde deben ubicarse varios servicios (cada uno con sus propiedades) en un número determinado de *clouds* (con sus limitaciones). Una vez definido el modelo, se deben comprobar los costes de todas las ubicaciones posibles, de modo que la mejor de ellas será la que presente un coste total inferior.

Para realizar una propuesta del modelo de coste se toman como referencia los cálculos de los costes fijos, variables y de transferencia de datos de [3] y el cálculo de los costes de computación de [9]. Aunque en [3] también se define el coste asociado al despliegue, el acceso a los datos requeridos para este cálculo es difícilmente conseguible en *clouds* públicos o privados gestionados por terceros, por la propia confidencialidad de las empresas.

A. Costes fijos

Se definen los siguientes parámetros para calcular los costes fijos, ya que se considera que estos son los más relevantes para estimar la inversión inicial realizada que permita confeccionar el *cloud* privado:

- C^{SeNe} : Coste de los servidores y los dispositivos de red utilizados.
- C^{SML} : Coste de las licencias software y middleware utilizadas.
- C^{Fa} : Coste de la instalación (infraestructura).
- C^{NeeCa} : Coste del equipamiento no electrónico (mobiliario, climatización, etc.) y del cableado.
- C^{Int} : Coste de la conexión a internet por mes.
- D_{IT} : Tiempo de depreciación de los dispositivos IT (36 meses).
- D_U : Tiempo de depreciación del equipamiento no electrónico y del cableado (120 meses).
- D_F : Tiempo de depreciación de la instalación (600 meses).
- M : Número de meses del cual se quiere calcular el coste fijo.

La fórmula para calcular el coste fijo es:

$$CostesFijos_M = \frac{M}{D_{IT}} (C^{SeNe} + C^{SML}) + \frac{M}{D_F} C^{Fa} + \frac{M}{D_U} C^{NeeCa} + C^{Int} \quad (1)$$

B. Costes variables

Se definen los siguientes términos para calcular los costes variables, ya que se consideran que estos son los necesarios

para estimar los costes que engloban todo lo necesario para que el *cloud* privado esté funcionando:

- C^{Elec} : Coste de la electricidad por unidad. Este valor se considera fijo.
- E_T^{Use} : Cantidad de electricidad consumida en el mes T.
- C_T^M : Coste de mantenimiento en el mes T.
- C_T^{Sto} : Coste de almacenamiento de datos en *clouds* en el mes T.

La fórmula para calcular el coste variable es:

$$CostesVariables_T = C_T^{Elec} \cdot E_T^{Use} + C_T^M + C_T^{Sto} \quad (2)$$

C. Costes de tráfico de datos

Se definen las siguientes variables para calcular los costes de tráfico de datos entre los diferentes servicios y los *clouds* utilizados. Estos costes dependerán de cuál sea la distribución de los servicios en los *clouds*:

- SP_T : Vector de ubicación de servicios en el mes T. Este vector tiene S coordenadas que se corresponde con el número de servicios. Por ejemplo, el vector (1,1,2) indica que los dos primeros servicios se ubican la instancia 1 y el tercer servicio en la instancia 2. El valor que puede tomar cada elemento del vector SP está comprendido entre 0 y P-1 donde P es el número total de instancias.
- DTM_T : Matriz del tráfico de datos. Es una matriz cuadrada de SxS dimensiones, donde la posición a_{ij} de la matriz representa la cantidad de datos transferidos (en GB) entre el servicio i-ésimo y el servicio j-ésimo durante el mes T.
- TCR : Matriz del coste de tráfico. Es una matriz cuadrada de PxP dimensiones, donde la posición β_{rt} de la matriz indica el coste de transferencia de un GB de datos desde la instancia r-ésima hacia la instancia t-ésima.

La fórmula para calcular el coste de tráfico de datos es:

$$CostesDeTráfico_T = \sum_{i=0}^{S-1} \sum_{j=0}^{S-1} DTM_T(i, j) \cdot TCR[SP_T(i), SP_T(j)] \quad (3)$$

D. Costes de computación

El cálculo de este ha modificado respecto a [9], ya que los datos requeridos para éste cálculo pues la definición propuesta se ajusta mejor a los datos que se pueden obtener del escenario que se plantea más adelante. Se definen las siguientes variables para calcular los costes de computación:

- VMQ_T : Vector de N elementos correspondientes al número de instancias diferentes. Cada posición indica el número de máquinas virtuales (instancias) de aquel tipo que se utilizan en el mes T.
- VMP : Vector de N donde cada posición indica el precio de cada una de las instancias.

Como se puede observar, los valores del vector VMQ tienen una relación directa con el vector SP, ya que estos valores cambian según la ubicación de los servicios. Por ejemplo, si el vector SP es (1,1,0) significa que se está utilizando la instancia 1 y la 0. En cambio, si el vector SP es (2,0,1) significa que se está utilizando las instancias 0, 1 y 2.

La fórmula para calcular el coste de computación es:

$$CostesComputación_T = \sum_{i=0}^{N-1} VMQ_T(i) \cdot VMP(i) \quad (4)$$

E. Coste total

El coste total será la suma de los costes anteriores:

- CF_M : Costes fijos de los M meses.
- CV_T : Costes variables en el mes T.
- CT_T : Costes de tráfico en el mes T.
- CC_T : Costes de computación en el mes T.

$$CosteTotal = CF_M + \sum_{T=1}^M CV_T + CT_T + CC_T \quad (5)$$

IV. ALGORITMO CMBSPP

Como se ha dicho anteriormente, para obtener la mejor ubicación de servicios posible es necesario calcular el coste para cada una de las combinaciones del vector SP. Se puede observar que este número total de combinaciones será igual a S^P , siguiendo el método *Branch&Bound* [16]. Se propone pues una herramienta software que realice el cálculo del coste para todas las combinaciones posibles del vector SP de manera automatizada y muestre cuál es la mejor de todas las combinaciones. El algoritmo que implementará esta herramienta (Fig. 2) y que ha sido llamado *Cost Model Based Service Placement Finder* (CMBSPP). En este algoritmo, puesto que los costes fijos y variables no dependen del vector SP, se propone calcularlos sólo una vez y no incluir su cálculo, ya que éste solo necesitará calcular los costes de tráfico y de computación para saber cuál es la mejor opción.

```

main(){
    configuraElEscenario();
    costeMinimo = maxRealNumber;
    while
    (AlgunaCombinacionDelVectorSPDiferentePosible){
        VectorSPActual=generaUnaCombinacionDiferente();
        calculaElVectorVMQ(VectorSPActual);
        calculaElCosteDeTráfico();
        calculaElCosteDeComputacion();
        coste=costeDeTráfico + costeDeComputacion;
        if (coste<costeMinimo){
            costeMinimo=coste;
            mejorVectorSP=VectorSPActual;
        }
    }
    muestraMejorCombinacion(mejorVectorSP, costeMinimo);
}
    
```

Fig. 2. Pseudocódigo del algoritmo CMBSPP

V. VALORACIÓN DEL ALGORITMO CMBSPP

Para evaluar si el modelo de coste propuesto y su algoritmo asociado ayudan a obtener la mejor ubicación de los servicios, se plantea un escenario a analizar donde la combinación con el coste económico menor será considerada la óptima. Con el propósito de comparar los resultados obtenidos y poder darles validez, se utiliza el *Free Cloud Cost Calculator* (FCCC) [17], una herramienta web que permite configurar un escenario con *clouds* públicos y estimar su coste. Más detalladamente, FCCC posee un amplio catálogo de instancias (*clouds* públicos) provistas por las principales empresas como Amazon, Microsoft, Google, Rackspace o HP. El usuario puede configurar qué instancias va a usar, durante cuánto tiempo y cuántos GBs de datos se van a transferir entre ellas y el *cloud* privado, entre otros

parámetros más complejos. Una vez configurado el escenario, FCCC calcula el coste de 3 años, detallando el coste de cada uno de los años. Esta herramienta ha sido creada por la compañía *RightScale* [18], empresa puntera en el ámbito del *cloud management*. Aunque es posible que la empresa utilice otro modelo de coste, desde este estudio se entiende que si los resultados obtenidos son similares o iguales, se puede dar validez al algoritmo CMBSPPF, ya que habrá llegado a la misma solución que un producto que se está comercializando. La descripción del escenario planteado, a modo de ejemplo, se puede apreciar en las Tablas I y II y en la Figura 3.

Se ha propuesto este escenario ejemplo en el que se ubican tres servicios que piden los mismos recursos pero que la transferencia de datos es distinta para comprobar la importancia de este factor. También se ha diseñado un *cloud* híbrido en el que no caben todos los servicios en el *cloud*

privado, ya que en este caso la solución sería evidente; y además se han escogido dos máquinas virtuales de distintas compañías que ofrezcan prestaciones similares para que la ubicación tampoco sea trivial a simple vista, ya que si fueran muy distintas, con sólo 3 servicios a ubicar se podría intuir la mejor opción. Se ha planteado un escenario con sólo tres instancias y tres servicios porque en FCCC se debe introducir el escenario manualmente cada vez que se quiere mirar una combinación del vector SP, de forma que si se tuvieran más instancias o servicios, el uso de esta herramienta sería más incómodo debido al número de horas/hombre si se quiere configurar todo el escenario si se pretende comprobar todas las combinaciones posibles. Los costes del *cloud* privado se consideran nulos porque se supone que ya se ha realizado la inversión para éste previamente.

Tabla I
ESCENARIO PLANTEADO - INSTANCIAS

ID	Proveedor	Instancia	RAM	HDD	Precio	Datos de entrada	Datos de salida
Cloud 0	Cloud Privado	VM propia	1,75GB	150GB	0€/h	0€/GB	0€/GB
Cloud 1	Azure	VM Role Small (A1)	1,75GB	70GB	0,085€/h (61,11€/mes)	0€/GB	0,0648€/GB
Cloud 2	EC2	m1.small	1,7GB	160GB	0,071€/h (50,92€/mes)	0€/GB	0,09€/GB

Tabla II
ESCENARIO PLANTEADO - SERVICIOS

ID	Consumo RAM	Consumo HDD	Transferencia de datos mensual
Servicio 0	0,75GB	20GB	15GB hacia Servicio 1 y 40GB hacia servicio 2
Servicio 1	0,75GB	20GB	15GB hacia Servicio 0 y 10GB hacia servicio 2
Servicio 2	0,75GB	20GB	10GB hacia Servicio 1 y 40GB hacia servicio 0

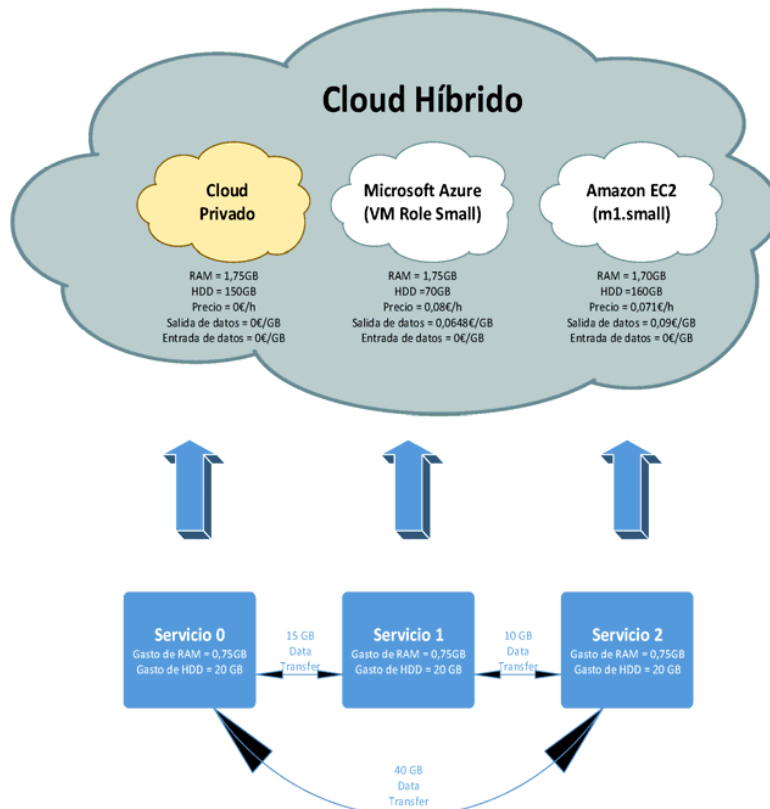


Fig. 3. Escenario planteado.

Se supone también que las máquinas virtuales que se utilicen estarán en funcionamiento durante todo el tiempo, debido a estar tratando un caso que requiere servicios de uso continuo y alta disponibilidad (p.e. gestión de datos de las Smart Grids de distribución). Los precios de Azure y EC2 han sido obtenidos de sus respectivas páginas web [19][20]. Los tráficos de datos entre los servicios definidos han sido considerados bidireccionales para facilitar la configuración del escenario en FCCC. Si se plantea calcular el coste de un mes entero, se obtienen las siguientes matrices del modelo de coste:

$$TCR = \begin{pmatrix} 0 & 0 & 0 \\ 0.0648 & 0 & 0.0648 \\ 0.09 & 0.09 & 0 \end{pmatrix} \quad (6)$$

$$DTM = \begin{pmatrix} 0 & 15 & 40 \\ 15 & 0 & 10 \\ 40 & 10 & 0 \end{pmatrix} \quad (7)$$

Como se ha comentado anteriormente, la limitación principal al utilizar el FCCC es que por cada combinación del vector SP se debe reintroducir el escenario completo. Por este motivo, primero se descartan las combinaciones del vector SP que no son posibles:

- SP = (1,X,X): Cualquier combinación que ubique el servicio 0 a Azure no será posible, ya que la máquina virtual que se utiliza tiene un espacio de 70 GB, mientras que el Servicio 0 necesita 20 GB más 55 GB que recibe de los otros servicios, sumando un total de 75 GB.
- SP = (0,0,0): Esta combinación no es posible porque el cloud privado tiene una capacidad de 1,75 GB de RAM y para tener los tres servicios ubicados allí serían necesarios 2,25 GB de RAM.
- SP = (2,2,2): Esta combinación no es posible por la misma razón que la anterior.
- SP = (0,1,1): Esta combinación no es posible porque ubicar los servicios 1 y 2 implicaría tener disponibles 20+20+40+15 = 95GB en Azure y solo se disponen de 70 GB.
- SP = (2,1,1): Esta combinación no es posible por la misma razón que la anterior.

Aunque se descartaron todas estas combinaciones de antemano (de las 27 totales sólo hay 14 posibles), se tardó alrededor de hora y media para configurar todos los escenarios y conseguir los costes de cada uno de ellos. Los resultados obtenidos con FCCC se muestran en la Tabla III.

Tabla III
RESULTADOS FCCC

Vector SP	Coste
(0,0,1)	67,81€
(0,0,2)	51,78€
(0,1,0)	64,96€
(0,1,2)	116,75€
(0,2,0)	51,78€
(0,2,1)	119,58€
(0,2,2)	103,55€
(2,0,0)	51,78€
(2,0,1)	119,58€
(2,0,2)	103,55€
(2,1,0)	116,75€
(2,1,2)	168,49€
(2,2,0)	103,55€
(2,2,1)	171,34€

Tabla IV
RESULTADOS CMBSPPF

Vector SP	Coste
(0,0,1)	64,35€
(0,0,2)	55,33€
(0,1,0)	62,73€
(0,1,2)	118,06€
(0,2,0)	53,08€
(0,2,1)	117,43€
(0,2,2)	106,70€
(2,0,0)	55,78€
(2,0,1)	120,13€
(2,0,2)	104,00€
(2,1,0)	118,51€
(2,1,2)	166,73€
(2,2,0)	106,25€
(2,2,1)	170,60€

De los resultados de la Tabla III, se observa que las mejores combinaciones son la (0,0,2), la (0,2,0) y la (2,0,0) con un coste de 51,78€, mientras que la peor de ellas es la (2,2,1) con un coste de 171,34€. En la Tabla IV se pueden apreciar los resultados de la misma prueba con el algoritmo CMBSPPF. Los resultados se han obtenido mediante la implementación del algoritmo en lenguaje Java y leyendo el escenario planteado desde un fichero de texto en formato JSON. Para encontrar la solución con este método se tardaron menos de 5 minutos en escribir el fichero JSON del escenario a analizar y a continuación sólo se tuvo que ejecutar el algoritmo, que obtuvo los resultados de forma instantánea.

Esta implementación del algoritmo permite una fácil integración con el resto de sistema del proyecto [21]. Cabe recordar que es el propio algoritmo el que va cambiando el vector SP (a diferencia de FCCC, donde se debe reconfigurar el escenario por cada combinación del vector SP) y también descarta por sí mismo las combinaciones que no son posibles. Como se puede observar, utilizando el algoritmo CMBSPPF se ha necesitado un tiempo considerablemente inferior para obtener los resultados. En este caso la mejor combinación es la (0,2,0) con un coste de 53,08€ y la peor es la (2,2,1) con un coste de 170,6€, obteniendo así la misma solución que con la herramienta FCCC. Si se ordenan todas las combinaciones posibles según su coste, se observa que el resultado es muy similar en los dos casos, excepto en las posiciones 9, 10 y 11 (ver Tabla V). También cabe destacar que las diferencias de coste entre los dos métodos son mínimas.

Tabla V
COMPARACIÓN DE LOS RESULTADOS OBTENIDOS

Orden	Vector SP	Coste CMBSPPF	Coste FCCC
1	(0,2,0)	53,08€	51,78€
2	(0,0,2)	55,33€	51,78€
3	(2,0,0)	55,78€	51,78€
4	(0,1,0)	62,73€	64,96€
5	(0,0,1)	64,35€	67,81€
6	(2,0,2)	104,00€	103,55€
7	(2,2,0)	106,25€	103,55€
8	(0,2,2)	106,70€	103,55€
9*	(0,2,1)	117,43€	119,58€
10*	(0,1,2)	118,06€	116,75€
11*	(2,1,0)	118,51€	116,75€
12	(2,0,1)	120,13€	119,58€
13	(2,1,2)	166,73€	168,49€
14	(2,2,1)	170,60€	171,34€

Estas diferencias en los precios pueden ser debidas a que FCCC utiliza un coste de tráfico para Amazon EC2 de 0€ y por esta razón algunas opciones obtienen el mismo coste cuando se ubica un servicio en este *cloud*. Este factor no permite diferenciar la mejor combinación entre ellas, pero en cambio sí es posible con CMBSPP. También existe una mínima diferencia entre el precio mensual de las instancias utilizado por FCCC y el de las webs de Azure y EC2.

VI. CONCLUSIONES

En este estudio se han realizado dos aportaciones importantes. La primera de ellas es una tabla de evaluación cualitativa de múltiples métricas que caracterizan un *cloud*. Esta tabla, contrastada entre diferentes grupos de investigadores y administradores de sistemas *cloud*, puede dar unas pautas de decisión sobre qué métrica/s utilizar y qué estrategia seguir para el estudio de ubicación de recursos, según la aplicación que se quiera llevar a cabo. La otra contribución de relevancia es la definición del algoritmo CMBSPP y su implementación en Java, que ha permitido la comparación con FCCC y que se integrará fácilmente en otras plataformas. La implementación del algoritmo CMBSPP proporciona una herramienta en la cual sólo es necesario describir una vez el escenario en un fichero de texto en formato JSON. La opción de crear una herramienta que remotamente utilizara FCCC arrastraría la misma problemática de reconfiguración del escenario para cada vector SP difícilmente automatizable y dependiente del mantenimiento del servicio por parte del proveedor externo.

Con esta comparación se ha validado parcialmente la bondad del modelo propuesto y su algoritmo asociado, ya que se obtienen los mismos resultados en cuanto a mejor y peor distribuciones y los precios obtenidos también son similares a los que ofrece una compañía puntera en el sector y que comercializa su herramienta. Sin embargo, aunque las pruebas realizadas en los diferentes escenarios de complejidad similar a la que aquí se ha expuesto proporcionan resultados similares, aún no se ha podido realizar una comparación exhaustiva planteando escenarios más complejos en número de instancias y número de servicios a ubicar, hecho que podría haber reafirmado la bondad del modelo propuesto. Dicha comparación queda para una posterior publicación aunque los resultados provisionales son prometedores. Como apunte final, se puede reforzar que una ubicación híbrida de los servicios, teniendo en cuenta los resultados obtenidos en el caso de estudio, será generalmente mejor que una únicamente pública, ya que las dos peores combinaciones del vector SP en ambos modelos no contemplan la utilización del *cloud* privado.

Por otro lado, se consideran una serie de líneas de investigación que continúen con este trabajo. En primer lugar, se deberían comparar los resultados de CMBSPP con los de FCCC utilizando un escenario que se considere lo suficientemente grande como para darle una validez definitiva al algoritmo y modelo de coste propuestos. Una vez validados, se debería probar la obtención de resultados a partir del algoritmo con escenarios cada vez más complejos, hasta llegar a obtener el límite que el algoritmo puede soportar para poder dar una solución en "tiempo real". Finalmente, se podría generar un *script* que fuera cambiando el fichero de configuración JSON en respuesta a escenarios variables en el tiempo de forma continua, de modo que el

algoritmo pudiera dar la mejor ubicación de recursos en este tipo de escenarios.

AGRADECIMIENTOS

Este trabajo se ha realizado dentro del marco definido por el proyecto europeo FINESCE, del 7th Framework Program (ICT-2011, grant number 604677).

REFERENCIAS

- [1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee et al. "A view of cloud computing." *Communications of the ACM* 53, no. 4, 50-58. 2010.
- [2] P. Mell, T. Grance. "The NIST definition of cloud computing.", 2011.
- [3] J. Altmann, "Cost model based service placement in federated hybrid clouds", *Future Generation Computer Systems*, vol. 41, pp. 79-90, 2014.
- [4] R. Van den Bossche, K. Vanmechelen, J. Broeckhove. "Cost-optimal scheduling in hybrid iaas clouds for deadline constrained workloads." In *Cloud Computing (CLOUD)*, 2010 IEEE 3rd International Conference on, pp. 228-235. IEEE, 2010.
- [5] F. Liu, J. Tong, J. Mao, R. Bohn, J. Messina, L. Badger, D. Leaf. "NIST cloud computing reference architecture." *NIST special publication* 500: 292. 2011.
- [6] Genez, Thiago AL, Luiz F. Bittencourt, and Edmundo RM Madeira. "On the performance-cost tradeoff for workflow scheduling in hybrid clouds." In *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing*, pp. 411-416. IEEE Computer Society, 2013.
- [7] M. Malawski, "Cost minimization for computational applications on hybrid cloud infrastructure", *Future Generation Computer Systems*, vol. 29, pp. 1786-1794, 2013.
- [8] H. Zhang, "Proactive Workload Management in Hybrid Cloud Computing", *IEEE Transactions On Network And Service Management*, vol. 11, n. 1, pp. 90-100, 2014.
- [9] O. Mazhelis, "Role of Data Communications in Hybrid Cloud Costs", En *37th EUROMICRO Conference on Software Engineering and Advanced Applications*, pp. 138-145, 2011.
- [10] V. K. Naik, "Workload Monitoring in Hybrid Clouds", En *IEEE Sixth International Conference on Cloud Computing*, pp. 816-822, 2013.
- [11] Z. Li, L. O'Brien, H. Zhang, R. Cai. "On a catalogue of metrics for evaluating commercial cloud services". In *Proceedings of the 2012 ACM/IEEE 13th International Conference on Grid Computing* (pp. 164-173). IEEE Computer Society, September, 2012.
- [12] Z. Li, L. O'Brien, R. Cai, H. Zhang, "Towards a taxonomy of performance evaluation of commercial Cloud services." In *Cloud Computing (CLOUD)*, 2012 IEEE 5th International Conference on (pp. 344-351). IEEE, June, 2012.
- [13] L. Wang, U.S. Khan. "Review of performance metrics for green data centers: a taxonomy study." *The Journal of Supercomputing*, 63(3), 639-656. 2013.
- [14] J.Y. Lee, J.W. Lee, S.D. Kim. "A quality model for evaluating software-as-a-service in cloud computing". In *Software Engineering Research, Management and Applications*, 2009. SERA'09. 7th ACIS International Conference on (pp. 261-266). IEEE, December, 2009.
- [15] Z. Li, L. O'Brien, H. Zhang, R. Cai. "Boosting Metrics for Cloud Services Evaluation--The Last Mile of Using Benchmark Suites." In *Advanced Information Networking and Applications (AINA)*, 2013 IEEE 27th International Conference on (pp. 381-388). IEEE, March, 2013.
- [16] P. Narendra, K. Fukunaga. "A branch and bound algorithm for feature subset selection." *Computers, IEEE Transactions on* 100, no. 9. 917-922. 1977.
- [17] Planforcloud website. [Disponible en línea: <http://www.planforcloud.com>, último acceso: 14/05/2015].
- [18] Rightscale website. [Disponible en línea: http://www.rightscale.com/home-v2?utm_expid=99127306-72_71-ToQdQIGVg6Yu9Y4vHQ.1, último acceso: 14/05/2015].
- [19] Amazon EC2 Virtual Machines prices. [Disponible en línea: <http://aws.amazon.com/es/ec2/pricing>, último acceso: 14/05/2015].
- [20] Microsoft Azure Virtual Machines prices. [Disponible en línea: <http://azure.microsoft.com/es-es/pricing/details/virtual-machines>, último acceso: 14/05/2015].
- [21] FINESCE Project website. [Disponible en línea: <http://www.finesce.eu/>, último acceso: 14/05/2015].

An objective video quality metric for live video streaming over cloud systems: full and non reference approaches

Jaume Segura-García, Santiago Felici-Castell, Miguel García-Pineda,
Enrique Chirivella-Pérez, Juan Gutierrez Aguado, José M. Claver
Dpt Informàtica, Universitat de València, Burjassot, Spain

jsegura@uv.es, felici@uv.es, migarpi@uv.es, Echirivella@gmail.com, jgutier@uv.es, jclaver@uv.es

Resumen—The increased adoption of smart phones, the access to mobile broadband networks and the availability of public Clouds allow new multimedia services, called Cloud Mobile Media. Nevertheless new challenges appear due to the higher and variable delay as result of the virtualization mechanisms in the Cloud. In particular, for live video streaming applications different considerations should be taken to keep a good Quality of Experience of the video in terms of a high Mean Opinion Score (MOS). We have performed different video transmissions using our own cloud infrastructure, measuring variables from the whole system. Our goal is to apply over these variables a statistical method, called Factor Analysis, to estimate the MOS both in Full Reference and Non Reference approaches, with two different points of view: the customer and the service provider. We compare the performance of the estimated MOS against publicly available video quality algorithms.

Palabras Clave—live video streaming, cloud computing, quality of experience, objective video quality metric, factor analysis

I. INTRODUCTION

Big companies like Google, Cisco Systems, Apple and Microsoft predict that by 2018, more than 90% of Internet traffic will be multimedia content (images, 3D images, ultra high definition video and audio, etc.) [1]. In addition, three recent developments: a) increased adoption of smart phones and tablets, b) increased access to mobile broadband networks, and c) availability of public Clouds are aligning to possibly enable a new generation of truly ubiquitous multimedia services on mobile devices, called Cloud Mobile Media (CMM) services [2] [3]. These infrastructures are replacing the traditional Content Delivery Networks (CDN) [4] because they are expensive solutions and indeed are designed mainly for static content (not optimized for applications such as live video streaming) and fail when requiring intensive processing and when the content cannot be stored ahead of time, as well as other issues related to scalability, heterogeneity and usability [3].

In this new scenario, content providers strive to offer the best Quality of Experience (QoE) for customers linked to their different Quality of Service (QoS) solutions. Therefore, developing accurate, perceptual-based quality metrics is a key requirement for these CMM multimedia services. As QoE is purely related to end-users, we analyze the effect of both core networks and cloud networks as a whole. However, subjective testing is time-consuming, expensive and requires special assessment facilities to produce reliable and reproducible test results. Notice that new challenges appear related to the QoE

management [5] over these cloud based infrastructures, in particular with live video streaming, due to the additional delay and jitter introduced because the virtualization processes and increase of network distance between end-user and service.

As many factors are affecting QoE in live streaming video as a CMM service, our goal is to apply an statistical method called Factor Analysis (FA) [6] to reduce the whole set of measurable variables, measured throughout from the streaming server to the end user, and to find out which ones have influence on the subjective video quality to define an objective video quality metric to estimate or predict the Mean Opinion Score (MOS), denoted by \widehat{MOS} . For the estimation, we have analyzed different approaches: Full Reference (*FR*) and Non Reference (*NR*) [7]. In *FR* approaches require access to both the original (or reference) and received video, while in *NR* not. In addition, for the *NR* approach, we consider two different points of view, the customer (called NR_{cu}) and the service provider (called NR_{sp}). Notice that these approaches have been designed in a robust and reliable way, taking into account variables from the whole live video streaming system and not only those directly related with the video signal.

Finally, in order to evaluate the accuracy of the proposed metrics (\widehat{MOS}) for the different approaches, we have measured the QoE using traditional subjective MOS through surveys following the recommendations given by ITU-R (BT.500-13 [8], P.910 [9]) and have compared these metrics with well known publicly available video quality algorithms following the recommendations given by Video Quality Experts Group (VQEG) [10].

The rest of the paper is structured as follows. Section II discusses the related work. Section III characterizes the delay and jitter in cloud infrastructures. Section IV describes the methodology, both the statistical method based on FA and the subjective video quality measurements by video quality surveys. Section V explains the network infrastructure and test bench, defining the observable variables related with QoS, bit stream and basic video quality metrics. Section VI details the different expressions to estimate the \widehat{MOS} for different approaches. Section VII analyzes and compares the performance of our proposed metric against well known video quality metrics. Finally, Section VIII concludes the paper.

II. RELATED WORK

In the literature we can find several works related to live video streaming over cloud infrastructures. In [11] the

authors propose a solution for providing scalable, seamless, live video streams service using cloud infrastructures and software defined networks to provide alternative paths to meet with the user demands. In [12] it is proposed a cloud-based media processing platform for enabling elastic live broadcasting on the processing of a large number of streams to adapt the user requirements, trying to minimize the use of cloud resources without affecting the quality of streams. The authors in [13] propose an algorithm to allocate the machines to host servers in a cloud infrastructure, for load balancing and task scheduling (in particular for transcoding) to deliver live content to meet the user demand, trying to minimize the use of resources (number of servers) and costs. From previous [11] [12][13] references, we can see that the main research is focused on the performance of the cloud, but not in the end user itself. It is worth mentioning that in [14] although it does not address live video streaming, the authors show a content distribution algorithm within CDNs using cloud infrastructures, taking into account not only on QoS but also on the QoE in terms of MOS, estimated using neural networks.

In relation to video quality metrics, a performance comparison of different video metrics in a subjective manner from different distortions using H.264/Advanced Video Coding (AVC) are done in [15] and [16], following the recommendations given by [10] in a similar manner as we do in Section VII. In addition, the authors in [17] evaluate the different Full Reference and Reduced Reference quality metrics in the context of audio and video, concluding that statistically speaking there are currently no objective metrics available that can replace subjective quality assessments and in particular Peak Signal to Noise Ratio (PSNR) shows the worst results from the subjective point of view. In the same vein, the study presented in [7] evaluates the same video quality metrics, used in Section VII, concluding that the MultiScale Structural SIMilarity index (MSSSIM), the natural visual feature based Video Quality Metric (VQM) and the perceptual spatio-temporal frequency-domain based MOTion-based Video Integrity Evaluation (MOVIE) indexes are the most reliable in terms of subjective MOS values. In [18] is proposed a novel QoE real-time assessment through a neural network and QoE-driven adjustment scheme, taking into account the video attributes (content type, frame rate, resolution) and network parameters (packet size, loss rate and delay) to estimate the PSNR that is mapped to MOS. Nevertheless the accuracy of the mapping MOS-PSNR process is weak (it does not follow a perceptual criteria), as is being done by [19]. In [20] is proposed a method to analyze and classify a set of monitoring variables in a video streaming application (such as network parameters, video type, users' profiles and terminal features) using machine learning algorithms to estimate the QoE in terms of MOS, but its performance is not compared with other video quality metrics and some assumptions, in particular to know the video type before hand is not feasible in a real environment.

Thus, our contribution is the proposal of an objective video quality metric for live video streaming services over CMM, in order to estimate the subjective MOS of the delivered video.

Tabla I: Average Round Trip Time and jitter in ms for the three scenarios.

Scenario	RTT [ms]	Jitter [ms]
Physical-Physical	0.121	0.008
Physical-VM	0.339	0.012
VM-VM	0.536	0.032

III. CLOUD MOBILE MEDIA CHARACTERIZATION

According to [5] in CMM infrastructures for live video streaming applications, the main inconvenient arises due to the additional delay and jitter mainly introduced by the virtualization processes and increase of network distance between end-user and service. Thus, in order to validate this statement, we have performed several experiments to measure the delay and jitter in three different scenarios as shown in Fig. 1. The testbed has been deployed in four physical nodes. Each node has the following hardware: 32 GB RAM, 2 Intel XEON CPU E5-2630 v3 @ 2.40 GHz with 8 cores with hyper threading, 2 TB Hard Disk, and 1Gbps NICs. In all physical nodes was installed Ubuntu 14.04 Server (Trusty Tahr).

Notice that the most noteworthy virtual bridging solutions mainly used inside the cloud infrastructures adopted by hypervisors (or Virtual Machine (VM) manager [21]) running on Linux systems are Linux Bridge (LB), the native kernel bridging module [22], and Open vSwitch (OVS), a distributed OpenFlow-enabled software-based switching facility capable of reaching kernel-level performance [23].

Thus, the scenarios selected are as followed. In the first scenario, we have measured these properties between two physical machines connected via a physical switch Cisco Catalyst WS-C2960-24TC-L (Fig. 1-a). In the second scenario, the communication is between a physical machine and a VM (Fig. 1-b). The VM is connected to a Linux bridge and this bridge is connected to the Ethernet interface. In the last scenario (Fig. 1-c), the communications are done between two VMs in a real deployment of OpenStack [24]. The VM is connected to a LB, this LB is connected to a OVS switch (known as integration bridge), that is finally connected to another OVS switch attached to the physical interface. The hypervisor used in all cases is KVM [25] and it has been optimized to use `vhost-net` networking mode, that bypasses the hypervisor and directly connects with the kernel.

We have used `ping` to measure the Round Trip Time (RTT) and `iperf` to obtain the jitter in all the scenarios. Table I summarizes these results. As we can see, the delay and the jitter increase as more virtualization layers are added, even four times more approximately. Notice that the values are small because all the machines (virtual and physical) belong to the same infrastructure (intracloud) and there is not extra back ground traffic.

IV. METHODOLOGY: STATISTICAL MODELING AND FACTOR ANALYSIS

Factor analysis is a statistical method used to reveal the full structure of the collected data and to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors [6],

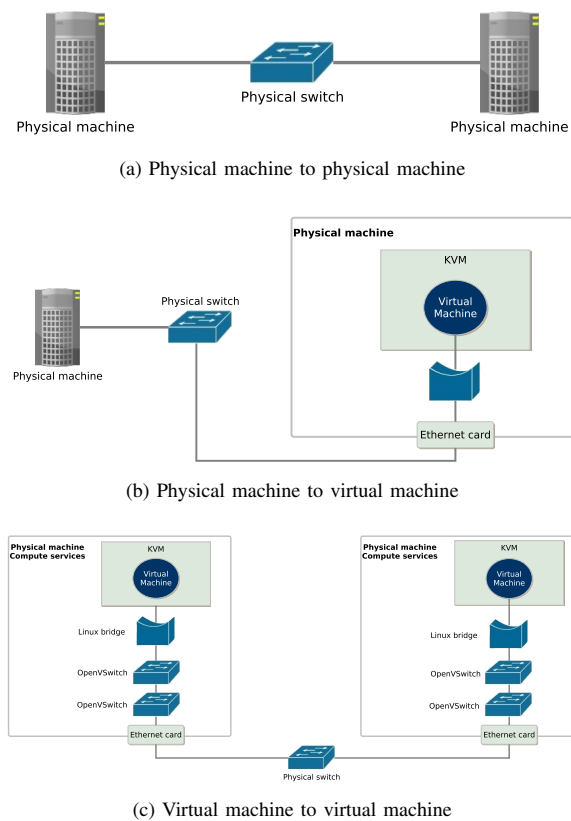


Fig. 1: Three different scenarios to measure the delay and jitter. a) Physical machine to physical machine. b) Physical machine to virtual machine. c) Virtual machine to virtual machine.

denoted by F_i where i is the factor identifier. The interdependencies between observed variables are used to define these factors. These factors are defined as a linear combination of a reduced set of observable variables, which partly explain the behavior of global variance of the variables. Using these factors, we can estimate the subjective MOS or \widehat{MOS} , as follows:

$$F_i = constant_i + \sum_{j=1} \alpha_j \cdot variable_j;$$

$$\widehat{MOS} = f(\dots, F_i, \dots); \quad i = 1, 2, \dots, n$$

where α_j are the coefficients of the linear regression for $variable_j$ in factor F_i and $f()$ is the regression function (using both linear and non-linear regressions) that estimates MOS. Notice that we define \widehat{MOS} using three different approaches: FR , NR_{cu} and NR_{sp} . In a simplified manner, in FA the extraction of the principal components is done by analyzing the correlation matrix of the different variables. Once we have a reduced set of components, we apply a rotation of these factors to obtain an orthogonal set by means of the *Varimax* algorithm and the *Kaiser* normalization [26]. The detail for the FA with different approaches is given in Section VI.

In order to perform the regression for the \widehat{MOS} expression, it is necessary to get at the same time real MOS values for

the different received videos in live video streaming. From the subjective tests defined at ITU-R BT.500-13 [8] and P.910 [9] we chose the Absolute Category Rating (ACR) that is a single stimulus method with both the original reference and the processed video being shown to the evaluators in a random sequence. The evaluators are unaware of its presence or its location in the displayed video set. The evaluators provide one rating for the overall video quality using a discrete five-level scale ranging from Bad (1) to Excellent (5). The study was conducted over two sessions, each lasting less than half an hour, as per recommendations in [8] in order to minimize evaluator fatigue. The evaluators' pool consisted of 35 undergraduate students of different ages (in average 21 years old) from our university. They are students (male and female) with a male majority, of the last course of Multimedia Engineering and they have enough knowledge about multimedia streaming.

Finally, following the recommendations given by VQEG [10], we compare the performance of the proposed objective quality metrics \widehat{MOS} for different approaches (FR , NR_{cu} and NR_{sp}) against other well known publicly available video quality algorithms. The detail of the performance evaluation is given in Section VII.

V. TEST BED: LIVE VIDEO STREAMING OVER CLOUD INFRASTRUCTURES

In order to evaluate the live video streaming as a CMM service, we have implemented a video client and a streaming server in different VMs, using the third scenario described in Section III. Besides for the wireless access, according to [27], we have selected a Guaranteed Bit Rate scenario, QoS Class Identifier (QCI=4). Overall, we have a mean end to end delay of 300 ms and 10% of jitter and a packet loss rate of 10^{-6} , following the arguments given in [5] and [27]. To adjust these parameters we use *iproute2* Linux utilities, in order to analyze the effect of both the core, cloud and access network as a whole.

The tool used to stream the video has been VideoTester [29] that allows video transmission with RTP over UDP unicast. VideoTester is furnished with a rich set of offline analysis capabilities, allowing us to extract the different variables, as shown in Table II, related to the QoS, bit stream and basic video quality parameters. All these variables are used in Section VI for the FA method. In order to emulate the live video streaming, the client requests a precoded video stream.

For the video transmissions, a set of 12 video sequences, spanning a wide range of content types in terms of video mobility, type of images, colours, etc. in order to generalize the different video streams, to design a MOS model (\widehat{MOS}) independent of the type of video, where 6 videos have been used to define the proposed model (called *akiyo*, *bride-far*, *football*, *foreman*, *mobile* and *news*) and the other 6 videos have been used for testing purpose to evaluate \widehat{MOS} (called *coastguard*, *flower*, *hall*, *paris*, *silent* and *tempete*). Fig. 2 shows a frame from each video. All these videos are available at [30].

These videos have a CIF (Common Intermediate Format, 352x288) resolution and have been coded with 3 different bit rates according to the different scenarios (256 kbps, 512 kbps and 768 kbps) using the codec H.264/AVC with the following coding options: 25 frames per second, profile Main@L1.3

Tabla II: Explanation of the measured variables related with QoS, bitstream and video quality parameters.

Name [unit]	Description
QoS	
Latency (L) [ms]	End to end delay
Delta (D) [ms]	Packet inter-arrival time measured per RTP packet
Jitter (J) [ms]	Delay variation
Skew (S) [ms]	Clock skew between sender and receiver from RTP timestamp
BandWidth (BW) [kbps]	Binary rate of video
Bit stream	
PLR [%]	Packet Loss Rate, ratio between number of lost and sent packets
PLD	Packet Loss Distribution measured by PLR per time
StreamEye (SE) [Byte/frame]	Visualization of the video stream framing structure at the receiver end, probably impaired with frame losses
GOP [size]	Group of Pictures or frames with inter dependences for decoding
IFLR [%]	I-Frame Loss Rate between number of I-frames lost and sent
Video Quality	
PSNR [dB]	Peak Signal to Noise Ratio in Y component
SSIM [index/frame]	Structural Similarity Index in Y component that measures the image degradation as perceived changes in structural information [28]
ITU-T G.1070	Opinion model for video applications based on coefficients in video quality estimation function with respect to coding and packet-loss degradations
DIV [%]	Distortion in Interval, maximum percentage of frames with a PSNR worse than original

and a GOP size of 50 frames. Notice that although we have other coding options, such as Scalable Video Coding (SVC) and other adaptive video streaming techniques, H.264/AVC is still most currently used in Internet and has attracted a lot of attention from industry and has been adopted by various application standards.

VI. RESULTS: OBJECTIVE VIDEO QUALITY METRICS

As previously described in Section IV, now we go into detail to perform FA with the variables obtained. We have used *SPSS* [31] software package for statistical analysis and *Matlab* [32]. In order to analyze potential hidden relations of the measured variables, we compare the information contributed by the first two standardized moments of the measured variables, *mean* (μ) and *standard deviation* (σ) in the next subsections. The final expressions for factors and MOS estimation are summarized in Table III for the different approaches FR , NR_{cu} and NR_{sp} as we will see through this section. Notice that from the variables described in Table II, for the FR approach we work with all the variables (14 variables in total), for the NR_{cu} approach we work only with the variables related to QoS and Bit Stream (10 variables in total) that can be monitored at the customer site, and finally for

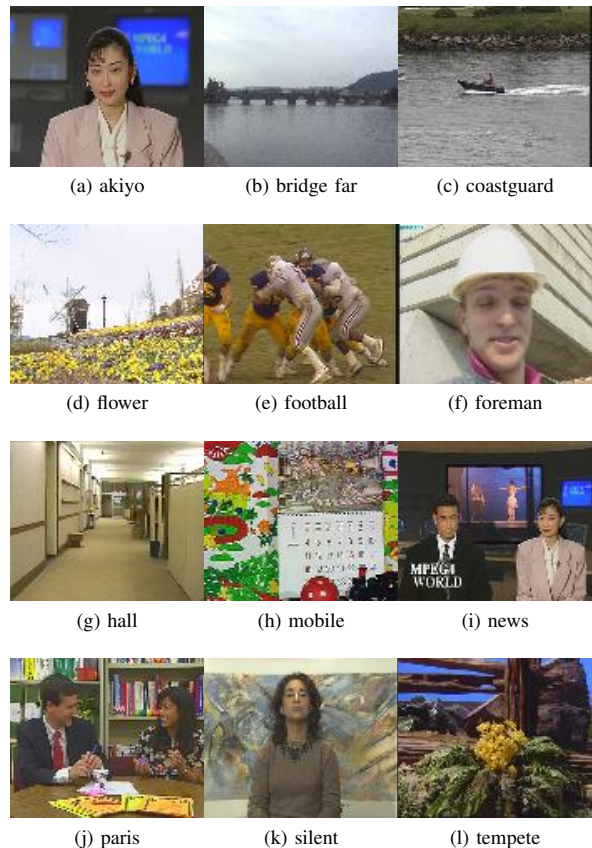


Fig. 2: Video sequences

the NR_{sp} approach we work only with the variables related with QoS, including PLD and PLR (7 variables in total) that can be monitored within the service provider.

A. Factor analysis with mean values

Applying FA to these sets of variables according to the different approaches FR , NR_{cu} and NR_{sp} , we obtain in all cases 3 factors, considering only eigenvalues greater than 1, explaining up to 84.7%, 91.58% and 94.46% of the total statistical variance respectively. Notice that the variance explained with NR_{cu} and NR_{sp} is higher than FR because we have a reduced set of variables. For the calculation of the factors, we apply the *Varimax* algorithm and *Kaiser* normalization [26] using a linear regression. For the subjective MOS expression based on the calculated factors, we use also a linear regression to estimate the real MOS, measured as explained in Section IV.

In Table IV, we compare the percentages of the average error in estimating the real subjective MOS (and its standard deviation) for the different approaches. The subjective assessment is explained in Section IV. From these results, we see that MOS_{FR} has an acceptable estimation error of 8.53 % with a standard deviation of 6.34, however for NR_{cu} and NR_{sp} are very high, around 30%. The factors and expression for FR are shown in Table III, with their respective squared Pearson's coefficients (R^2 [33]). To reduce the estimation error for NR_{cu} and NR_{sp} , in next subsection we will consider

Tabla III: Final expressions for factors and MOS estimation with different approaches.

FR	
\widehat{MOS}_{FR}	$3.166 + 0.01 \cdot F_1 - 0.17 \cdot F_2 + 0.9 \cdot F_3; (R^2 = 0.94)$
F_1	$-1.935 + 0.002 \cdot BW - 0.067 \cdot D + 0.456 \cdot G1070 + 0.0003 \cdot SE; (R^2 = 0.99)$
F_2	$-0.204 - 10.627 \cdot IFRL + 0.033 \cdot J; (R^2 = 0.96)$
F_3	$-45.151 + 0.042 \cdot PSNR - 0.01 \cdot DIV + 1.69 \cdot SSIM + 0.27 \cdot L - 0.0003 \cdot S; (R^2 = 0.99)$
NR _{cu}	
$\widehat{MOS}_{NR_{cu}}$	$2.039 - 0.18 \cdot F_1 + 1.124 \cdot F_2 + 0.685 \cdot F_3 - 0.054 \cdot F_1^2 + 1.125 \cdot F_2^2 + 0.123 \cdot F_3^2; (R^2 = 0.92)$
F_1	$-3.154 + 0.0004 \cdot BW + 0.000005 \cdot varBW + 0.09 \cdot D + 13498105 \cdot PLD + 15703088 \cdot PLR - 0.0001 \cdot SE - 5.99E - 8 \cdot varSE; (R^2 = 0.9)$
F_2	$-0.446 + 0.001 \cdot varD - 11 \cdot IFRL - 0.036 \cdot J + 0.021 \cdot varJ; (R^2 = 0.98)$
F_3	$-39.53 - 0.00008 \cdot S + 9.81E - 7 \cdot varS + 0.246 \cdot L; (R^2 = 0.97)$
NR _{sp}	
$\widehat{MOS}_{NR_{sp}}$	$2.02 - 0.006 \cdot F_1 - 0.503 \cdot F_2 + 0.364 \cdot F_3 + 0.024 \cdot F_1^2 + 0.68 \cdot F_2^2 + 0.522 \cdot F_3^2; (R^2 = 0.76)$
F_1	$1.45 + 0.001 \cdot BW + 0.00002 \cdot varBW - 0.065 \cdot D - 8297046 \cdot PLD - 8276808 \cdot PLR - 0.0004 \cdot SE + 7.82E - 8 \cdot varSE; (R^2 = 0.97)$
F_2	$-48.9 - 0.002 \cdot S - 5.209E - 7 \cdot varS + 0.302 \cdot L; (R^2 = 0.99)$
F_3	$-2.664 + 0.002 \cdot varD + 0.004 \cdot J + 0.046 \cdot varJ; (R^2 = 0.87)$

Tabla IV: Average errors in estimating real MOS in % (and the standard deviation) using mean values and linear regressions in the FA for different approaches.

scenario:	FR	NR _{cu}	NR _{sp}
% avg (std)	8.53 (6.34)	26.53 (13.22)	28.99 (10.64)

the use both of the mean and standard deviation of the variables, as well as non-linear regression methods.

B. Factor analysis with mean and variance

Following the same steps as explained in previous subsection, but now considering both the mean and variance of the variables for the NR_{cu} and NR_{sp} approaches, we obtain also 3 factors with eigenvalues greater than 1, that explain up to 84.54% and 85.49% respectively. Notice that the total variance explained now is lower than in subsection VI-A, because we are considering more variables. In these expressions the standard deviation for a variable X is denoted by $varX$. Using these factors, we calculate the subjective MOS expression with three different methods, using both a linear regression and two non-linear regressions. The non-linear regressions are done using both a quadratic and a cubic polynomial function.

We compare the percentages of the average error (and its standard deviation) in estimating the real subjective MOS for the NR_{cu} and NR_{sp} approaches with a linear, quadratic and cubic expression as shown in Table V. On the one hand, we see that the linear regression have an excessive error (above 25%). On the other hand, using a quadratic polynomial

Tabla V: Average errors in estimating real MOS in % (and the standard deviation) using mean and variance values using the linear, quadratic and cubic expressions for NR_{cu} and NR_{sp} approaches.

scenario:	NR _{cu}	NR _{sp}
(% avg (std))		
Linear	27.35 (28.97)	27.77 (25.20)
Quadratic	15.80 (11.42)	15.94 (13.85)
Cubic	28.29 (18.19)	20.86 (12.64)

function in both approaches show good results, an average error of 15.80% and 15.94% for NR_{cu} and NR_{sp} respectively (although the accuracy given by NR_{cu} is greater than NR_{sp} as we would expect because the standard deviation is lower), while the estimating error using the cubic expression is high, more than 20%. Thus, after seeing these results, the factors and the \widehat{MOS} expressions for NR_{cu} and NR_{sp} using mean and variance with quadratic regressions are shown in Table III, with their respective squared Pearson's coefficients.

VII. PERFORMANCE EVALUATION

After the previous discussion, in order to evaluate the accuracy of estimated subjective MOS using the expressions for FR, NR_{cu} and NR_{sp} as shown in Table III, we follow the recommendations given by VQEG [10]. This process consists of adjusting the scatter plot of the estimated MOS vs. measured MOS values, using in this case a nonlinear monotonic regression with a 4-parameter cubic polynomial fitted to the data with the best fit in a least squares sense. The functional form of the nonlinear regression is not critical as specified in [10].

Figs. 3a-3c show the nonlinear regression and the scatter plots for our proposed models using the test videos. In the same manner, Figs. 3d-3g show also the scatter plots and the nonlinear regression (using the same video sequences) for well known publicly available objective video quality algorithms, described in Table VI. These video quality metrics are implemented in the MSU video measurement tool [34]. Once we have plotted the objective metrics vs the subjective MOS, we perform the statistical analysis, described in [10], to show the evaluation metrics relating to: a) prediction of accuracy, by the Pearson Correlation Coefficient (PCC or R [33]) and Root Mean Square Error (RMSE), b) monotonicity by Spearman Rank Order Correlation coefficient (SROCC), and c) consistency (Outlier Ratio). More detail for these evaluation metrics can be found in [7]. The fidelity of an objective quality assessment metric to the subjective assessment is considered high if PCC and SROCC are close to 1, as well as OR and RMSE are low, close to 0. As shown in Table VII, on the one hand, we can see that PCC and SROCC for all the metrics, except for \widehat{MOS} in NR_{sp} and MSSSIM, are above 0.93, which are good results for video, and on the other hand RMSE and OR are smaller than 0.3 for for all the metrics, except for \widehat{MOS} in NR_{sp}, which means that most of the MOS computed by the metrics are within the grades given by the evaluators. It is worth mentioning that PSNR, although it does not follow the human visual criteria, it has a good performance due to the packet loss rate of 10^{-6} , as detailed in Section V and

Tabla VI: Objective video quality metrics.

Metric	Description
PSNR	Peak Signal-to-Noise Ratio is a numerical measurement based on the mean square error calculated for each frame (Y component only)
VQM [35]	Video Quality Metric is DCT-based metric which exploits the property of visual perception and it is contained in ITU-T recommendation J.144 [36]
3SSIM [37]	3SSIM is based on region division (edges, textures and smooth regions) of source frames. Notice that the human eye can see difference on textured or edge regions precisely than on smooth regions. The result metric is calculated as weighted average of Structural SIMilarity (SSIM) metric for those regions.
MSSSIM [38]	MultiScale SSIM based on SSIM metric of several downscaled levels of original images. Result is weighted average of those metrics. MSSSIM accounts for the multiscale nature of both natural images and human visual system

Tabla VII: Evaluation metrics to compare the performance of the different (\widehat{MOS}) expressions with different approaches against other video quality algorithms.

algorithm:	PCC	SROCC	RMSE	OR
\widehat{MOS}_{FR}	0.98	0.89	0.19	0.27
$\widehat{MOS}_{NR_{cu}}$	0.97	0.93	0.29	0.28
$\widehat{MOS}_{NR_{sp}}$	0.86	0.83	0.69	0.11
VQM	-0.98	-0.96	0.15	0.20
3SSIM	0.94	0.94	0.26	0.06
MSSSIM	0.92	0.91	0.30	0.09
PSNR	0.99	0.97	0.12	0.07

the proposed metric for \widehat{MOS} in NR_{sp} is not too accurate due to the lack of information and variables introduced in the model. The last statement is also confirmed by the high standard deviation as shown in Fig. 3c.

VIII. CONCLUSION

The increasing adoption of smart phones and tablets, the increased access to mobile broadband networks, and the availability of public Clouds allow a new kind of services, called Cloud Mobile Media services. Nevertheless applications such as live video streaming over these infrastructures, should take into account a good subjective quality of the delivered video, or QoE in terms of perceptual subjective MOS, due to the rapid changing network conditions, the increased delay and jitter and the limited resources of the mobile devices. Because many variables affect the subjective MOS, in this paper we have used a statistical method called Factor Analysis, considering all kind of variables related to QoS, bitstream and basic video quality metrics. With the analysis we have found different expressions for each approach FR , NR_{cu} and NR_{sp} , to estimate the MOS (\widehat{MOS}) with a mean estimation error of 8.53%, 15.80% and 15.94% respectively.

As future work, the proposed approach could be extended including additional variables both from the Service Provider (e.g. inner variables to the system) and the customer (e.g. resources of the mobile devices) point of view, as well as considering just noticeable differences on these variables. Also, we could consider different time scales of the variables to improve our model.

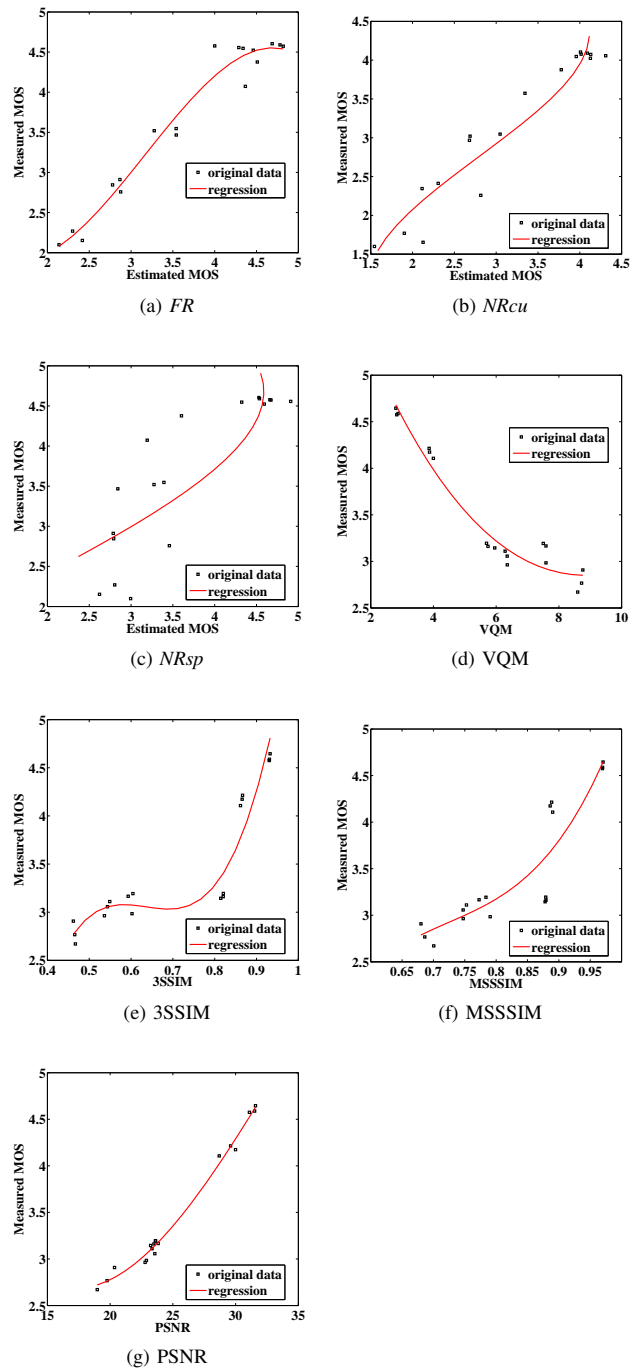


Fig. 3: Scatter plots: objective metrics vs. measured MOS

ACKNOWLEDGMENTS

This work was supported by the Universitat de València under the project UV-INV-PRECOMP14-207134 and by the Spanish Ministry of Economy and Competitiveness and FEDER under project TEC2012-37945-C02-02

REFERENCIAS

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2013-2018. White paper." Cisco Systems, Corp., Tech. Rep., 08 2014. [Online]. Available:

- http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf
- [2] S. Dey, "Cloud Mobile Media: Opportunities, challenges, and directions," in *Computing, Networking and Communications (ICNC), 2012 International Conference on*, Jan 2012, pp. 929–933.
 - [3] Y. Wen, X. Zhu, J. Rodrigues, and C. W. Chen, "Cloud Mobile Media: Reflections and Outlook," *Multimedia, IEEE Transactions on*, vol. 16, no. 4, pp. 885–902, June 2014.
 - [4] G. Pallis and A. Vakali, "Insight and Perspectives for Content Delivery Networks," *Commun. ACM*, vol. 49, no. 1, pp. 101–106, Jan. 2006. [Online]. Available: <http://doi.acm.org/10.1145/1107458.1107462>
 - [5] T. Hobfeld, R. Schatz, M. Varela, and C. Timmerer, "Challenges of qoe management for cloud applications," *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 28–36, April 2012.
 - [6] R. L. Gorsuch, *Factor analysis*. Hillsdale: Lawrence Erlbaum Associates, 1983.
 - [7] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 165–182, June 2011.
 - [8] "Methodology for the Subjective Assessment of the Quality of Television Pictures," *ITU-R, BT.500-13*, 1 2013. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-1!!PDF-E.pdf
 - [9] "Subjective Video Quality Assessment Methods for Multimedia Applications," *ITU-R, P.910*, 9 1999.
 - [10] "Final report from the video quality experts group on the validation of objective models of video quality assessment. Phase I," Video Quality Experts Group (VQEG), Tech. Rep., 06 2000. [Online]. Available: <http://www.its.bldrdoc.gov/vqeg/projects/rtv-phase-i/rtv-phase-i.aspx>
 - [11] Q. Wang, K. Xu, R. Izard, B. Kribbs, J. Porter, K.-C. Wang, A. Prakash, and P. Ramanathan, "GENI Cinema: An SDN-Assisted Scalable Live Video Streaming Service," in *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, Oct 2014, pp. 529–532.
 - [12] B. Cheng, "Mediapaas: A cloud-based media processing platform for elastic live broadcasting," in *Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on*, June 2014, pp. 713–720.
 - [13] J.-C. Gregoire and A. Hamel, "On scheduling live media streaming in the cloud: A study," in *A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2014 IEEE 15th International Symposium on*, June 2014, pp. 1–6.
 - [14] H. A. Tran, A. Mellouk, and S. Hoceini, "QoE Content Distribution Network for Cloud Architecture," in *Network Cloud Computing and Applications (NCCA), 2011 First International Symposium on*, Nov 2011, pp. 14–19.
 - [15] I. Sedano, M. Kihl, K. Brunnstrom, and A. Aurelius, "Evaluation of video quality metrics on transmission distortions in H.264 coded video," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2011 IEEE International Symposium on*, June 2011, pp. 1–5.
 - [16] A. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. Bovik, "Wireless Video Quality Assessment: A Study of Subjective Scores and Objective Algorithms," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 4, pp. 587–599, April 2010.
 - [17] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audiovisual services: A survey," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482 – 501, 2010, special Issue on Image and Video Quality Assessment. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0923596510000299>
 - [18] Y. Lan, Y. Geng, L. Rui, and A. Xiong, "Video quality assessment and QoE-driven adjustment scheme in wireless networks," in *Communication Technology (ICCT), 2012 IEEE 14th International Conference on*, Nov 2012, pp. 46–50.
 - [19] J. Gross, J. Klaue, H. Karl, and A. Wolisz, "Cross-layer optimization of OFDM transmission systems for MPEG-4 video streaming," *Computer Communications*, vol. 27, pp. 1044–1055, 2004.
 - [20] M. Mushtaq, B. Augustin, and A. Mellouk, "Empirical study based on machine learning approach to assess the QoS/QoE correlation," in *Networks and Optical Communications (NOC), 2012 17th European Conference on*, June 2012, pp. 1–7.
 - [21] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, 2013. [Online]. Available: <http://dx.doi.org/10.1002/wcm.1203>
 - [22] "The Linux Foundation, Linux Bridge," accessed: 15/4/2015. [Online]. Available: <http://www.linuxfoundation.org/collaborate/workgroups/networking/bridge>
 - [23] B. Pfaff, J. Pettit, T. Koponen, K. Amidon, M. Casado, and S. Shenker, "Extending networking into the virtualization layer," in *Proc. of workshop on Hot Topics in Networks (HotNets-VIII)*, 2009.
 - [24] "OpenStack: The Open Source Cloud Operating System. Juno release," accessed: 15/4/2015. [Online]. Available: <http://www.openstack.org/>
 - [25] "Kernel Based Virtual Machine," accessed: 15/4/2015. [Online]. Available: http://www.linux-kvm.org/page/Main_Page
 - [26] H. T. Kaiser, "The Application of Electronic Computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, p. 141151, 1960.
 - [27] M. Alasti, B. Neekzad, J. Hui, and R. Vannithamby, "Quality of service in WiMAX and LTE networks [Topics in Wireless Communications]," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 104–111, 2010.
 - [28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
 - [29] I. Ucar, J. Navarro-Ortiz, P. Ameigeiras, and J. M. Lopez-Soler, "Video Tester – A multiple-metric framework for video quality assessment over IP networks," *CoRR*, vol. abs/1301.5793, 2013.
 - [30] "Video Samples," accessed: 16/10/2014. [Online]. Available: <http://www2.tkn.tu-berlin.de/research/evalvid/cif.html>
 - [31] "Statistical Package for the Social Sciences (SPSS)," accessed: 15/10/2014. [Online]. Available: <http://www.ibm.com/software/en/analytics/spss/>
 - [32] "Matlab," accessed: 15/10/2014. [Online]. Available: <http://www.mathworks.com/products/matlab/>
 - [33] W. Härdle and L. Simar, *Applied multivariate statistical analysis*. Springer, 2007, vol. 22007.
 - [34] "MSU Video Quality Measurement Tool," accessed: 15/10/2014. [Online]. Available: http://compression.ru/video/quality_measure/video_measurement_tool_en.html
 - [35] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *Broadcasting, IEEE Transactions on*, vol. 50, no. 3, pp. 312–322, Sept 2004.
 - [36] "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference (superseded)," *ITU-R J.144*, 6 2001. [Online]. Available: <http://www.itu.int/rec/T-REC-J.144-200103-S/>
 - [37] C. Li and A. C. Bovik, "Content-weighted video quality assessment using a three-component image model," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 003–011 003–9, 2010. [Online]. Available: <http://dx.doi.org/10.1117/1.3267087>
 - [38] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, Nov 2003, pp. 1398–1402.

Evaluation of the Trade-off between Energy Consumption and QoE in Video Transmission with MANET Routing Protocols

Ramon Sanchez-Iborra, Maria-Dolores Cano

Departamento de Tecnologías de la Información y las Comunicaciones

Universidad Politécnica de Cartagena

Antiguo Cuartel de Antigonos. Plaza del Hospital, N° 1, 30202 Cartagena (Murcia)

{ramon.sanchez, mdolores.cano}@upct.es

Abstract- The popularity of video applications via streaming, conferencing, or sharing has put energy saving on the top of current demands for mobile and network devices. However, whereas end users enjoy a plethora of wireless devices for consuming the new wave of Internet services, batteries' lifetime has not been as extended as it would be desirable. For that reason, reducing power draining in every task performed by these devices is crucial, especially in networking tasks, which has an important weight in mobile devices' energy consumption. In this work, we present the performance evaluation of three remarkable ad-hoc routing protocols supporting a demanding video-streaming service in terms of energy consumption. The studied protocols are Better Approach To Mobile Adhoc Networking (BATMAN), Optimized Link State Routing (OLSR), and Ad hoc On-Demand Distance Vector (AODV). Furthermore, the quality attained in the multimedia transmissions is also considered from the user perspective (Quality of Experience, QoE). Thus, the trade-off between the QoE achieved by the video communication and the energy consumed by the routing protocols is examined. Additionally, we introduce an enhancement to turn the BATMAN routing protocol more energy-efficient. Results reveal that an increase on power draining of the routing tasks not necessarily leads to a QoE improvement, as it is possible to reduce the ad-hoc routing protocol's power consumption without harming (even increasing) the quality of the multimedia service flowing through the network.

Keywords- Energy consumption, QoE, video-streaming, MANET

I. INTRODUCTION

The development of power-efficient techniques to diminish energy consumption in communication networks is a compelling need. In addition to global warming, economical interests, and other large-scale issues, there are other small-scale reasons for reducing energy-consumption levels when consuming networking services. These reasons are mostly related to the usability of end-devices, e.g., tablets or cell-phones. Unlike the electronic technology, batteries' capacity does not follow the Moore's law [1], so extending gadgets lifetime is a greatly valued asset. Wireless cards energy consumption has a remarkable weight in mobile devices' power draining [2], thereby, it is necessary the development of networking protocols and procedures as efficient as possible in order to extend battery duration.

One of the wireless technologies receiving more attention from the research community is Mobile Ad-hoc NETWORKS (MANET). Under this scheme, nodes composing the system auto-configure themselves aiming at creating a mesh network without using a central point, e.g., a router, in charge of network-managing tasks. Thus, nodes forming the ad-hoc network help each other by relaying packets to their final destination. These networks usually employ the IEEE 802.11 (WiFi) standard and present highly attractive characteristics for the quick deployment of networks without using a pre-existent infrastructure. However, due to the dynamic nature of MANETs, efficient routing protocols are necessary in order to reach any node within the network.

Regarding the Internet-usage trend, it is clear that video-communications will be one of the most employed services in the upcoming years [3]. Nevertheless, this is one of the most demanding services in terms of network efficiency and reliability, as it is extremely low tolerant to packet loss and delay variations. Although in wired networks these issues can be maintained under control, the variable nature of wireless systems makes them much tougher to manage. Focusing in the case of MANETs, this fact represents another challenge to ad-hoc routing protocols, which not only have to find the most proper route to a given destination, but also should not interfere in the multimedia communication, avoiding harming the quality provided to the end-user. Regarding quality measurements, the current trend is estimating what the end user actually perceives when consuming the service. This paradigm is called Quality of user Experience (QoE) and differs from the traditional Quality of Service (QoS) approach in the human perspective given to quality estimations [4]. In order to measure the subjective opinion that a user has about a given service, different approaches can be taken. The most accurate methodology is performing subjective tests in which a panel of human testers rates the quality of the service under evaluation. This approach returns a quality score, so-called Mean Opinion Score (MOS), for the considered service in a scale 1 - 5. This methodology presents several drawbacks such as its high expenses and the impossibility of performing tests at real time. For those reasons, several mathematical models have been presented in order to provide an estimation of the QoE (MOS) reached by different multimedia services [5].

In this paper, we present a comparative study aiming at understanding the trade-off between the quality perceived when consuming a video-service and the energy consumed by the end-device. Concretely, we focus on measuring the energy consumed by wireless cards, ignoring other sources of battery draining such as video (de) coding or other processing tasks. To this end, three different routing protocols, namely BATMAN (Better Approach To Mobile Adhoc Networking) [6], OLSR (Optimized Link State Routing) [7], and AODV (Ad hoc On-Demand Distance Vector) [8] are evaluated while managing an ad-hoc network with different number of video-streaming transmissions flowing through it. BATMAN and OLSR are two of the most representative pro-active ad-hoc routing protocols, whereas AODV follows a reactive approach. This evaluation is performed under different environmental conditions and motion situations, evaluating for each protocol the attained QoE, in terms of MOS, for the video transmissions and the energy needed to accomplish this task.

After the evaluation of these standardized protocols, we propose an enhancement to the BATMAN algorithm, not just aiming at improving the quality of the video streaming service, but also trying to reduce the energy consumption of mobile devices. In a previous work [9], it was demonstrated the need of properly tuning the time-interval between BATMAN's OriGinator Message (OGM) broadcasts (hereinafter OGM interval) in order to adapt this routing protocol to current network conditions. Thus, a characterization of the dynamic update of the OGM interval is provided here, presenting a mathematical formulation that permits to fit this control-packets sending-interval to the network load in real-time. By means of this enhancement, the BATMAN routing protocol turns to be more flexible and suitable to better support different types of traffic. Through intensive computer simulation, the performance of BATMAN supporting a video-streaming service is evaluated, comparing the results attained by the original protocol with those obtained including the proposed dynamic adaptation of the OGM interval.

Thereby, the main contributions of this work are the following:

- An evaluation of the trade-off between energy consumption and QoE of the video-transmission service by using the aforementioned routing protocols.
- The design and evaluation of a novel proposal to enhance the BATMAN routing protocol in terms of energy consumption and QoE.

The rest of the paper is structured as follows. Section II presents an overview of the most relevant works studying the trade-off between QoE and energy consumption. The employed test-bench is thoroughly described in Section III. The attained results and their discussion compose Section IV. The paper ends summarizing the most important facts in Section V.

II. RELATED WORK

The popularity of video applications via streaming, conferencing, or sharing has put energy saving on the top of current demands for mobile and network devices. The Green Metadata standard [10], [11], currently in its final approval stage at ISO (International Organization for Standardization),

is a good example of the highly importance of power reduction in video transmission, and how it is possible to achieve it compromising QoE as less as possible from mobile devices perspective. In fact, exploring energy consumption versus QoE has produced its first results in several technologies and services such as Long Term Evolution, WiFi, or Wimax, among others [12]–[18].

Facing the QoE-and-energy aware video transmission from the physical layer, authors in [17] proposed to apply multicasting in an energy-efficient radio resource allocation scheme based on video quality as perceived by the end user. Their proposal was tested in an Orthogonal Frequency-Division Multiplexing (OFDM) scenario using a genetic algorithm as the basis of the allocation scheme. Results suggested that it is possible to steadily control energy requirements while maintaining high levels of perceptual quality as the number of subscribers increases in different multicasting scenarios. Nevertheless, the video quality method employed in this work, namely ITU-T Rec. J-144 [19], is outdated for video streaming in wireless environments and using the new ITU-T recommendations would be suggested to extend the validity of authors' proposal. Xu *et al.* [15] introduced two new performance metrics called QoE-aware Energy Efficiency (QEE) and QoE-aware Spectral Efficiency (QSE). Based on these parameters, which provide information about power consumption and bandwidth consumption from video quality's perspective, authors proposed a new spectrum and energy efficient mobile association and resource allocation scheme for wireless heterogeneous networks. In their work, video QoE is characterized in terms of Peak Signal to Noise Ratio (PSNR), and the solution proposed is formulate as an association and allocation optimization problem. After evaluating the system performance including PSNR distribution and resource utilization, they concluded that it greatly depends on the bandwidth and power decaying factors of cellular networks, leaving for further research the study of trade-off between QEE and QSE.

Signal *et al.* [12], [13] advanced the state of the art by presenting a cross-layer optimization framework to enhance QoE and energy efficiency of heterogeneous wireless multimedia broadcast end devices. Their solution combines user composition-aware source coding rate (Scalable Video Coding) optimization, optimal time-slicing for layer coded transmission, and a cross-layer adaptive modulation and coding scheme. Simulation results in a single-cell video broadcast network using Digital Video Broadcast-hand-held (DVB-H) technology revealed that it is possible to group users based on their device capabilities and estimated channel conditions experienced by them and to broadcast adaptive content to these groups, resulting in a better performance than current employed techniques.

Focusing on MANET, numerous works have studied and proposed energy-aware solutions for routing protocols (e.g., [20]–[22]). However, there are not related works to the authors' knowledge when it comes to evaluate both QoE and energy consumption. Only a few works can be found that aim to improve energy efficiency and QoS, i.e., quality from the classical network perspective, and most of them for on demand routing protocols [23]–[29]. As an example, Chen *et al.* [23] proposed a modification to the AODV protocol by introducing bandwidth estimation through the dissemination of bandwidth information in AODV hello messages. They

obtained satisfactory results, although no node-movement was included in their evaluated scenarios. Other similar works are able to improve delay, packet delivery ratio, or jitter, but 1) the formal evaluation of quality of user experience is not included, 2) many of them are focused only on stationary nodes, and 3) most of them do not cover video streaming. With our work, we aim to fill this gap providing a first approach to QoE-energy design of routing protocols in MANET for video delivery.

III. TEST BENCH

As aforementioned, the results presented in this work have been obtained by means of computer simulation. Specifically, the Omnet++ v.4.4 network simulator has been used, together with the InetManet 2.0 framework [30]. In the following, a detailed description of the deployed scenario and the employed tools is provided.

The network under consideration was an IEEE 802.11g wireless network at 54 Mbps in ad-hoc mode. Channel access was according to the Distributed Coordination Function (DCF) scheme without employing the RTS/CTS mechanism. It was configured a mesh network composed by 25 nodes and its performance was evaluated with a variable number (1, 3, and 5) of simultaneous video streams flowing through it. The nodes composing the network were spread within an area of 500 m x 500 m, thus having a node density of 100 nodes/km². Aiming at obtaining statistically consistent results, 10 independent simulation instances for each scenario were run, taking the average value for every measured parameter. It has been also calculated the standard deviation and the 95% confidence intervals for every measurement. Each simulation-run lasted 75 s.

Regarding the node's hardware configuration, it has been simulated the normal operation of a mobile handset, namely, a cell-phone. Thus, each node's wireless card has been configured following the real parameters of the widely used Broadcom's BCM4330 IEEE 802.11 a/b/g/n chip [31]. This chipset has been installed in different brand's cell-phones and tablets, namely, iPhone 4s, iPad 3, Samsung Galaxy S 2, or Samsung Galaxy Tab 2, among others. Thereby, the transmission power was set to 20 mW (13 dBm), with a sensitivity at 54 Mbps of -74 dBm and a SNIR (Signal to Interference and Noise Ratio) threshold of 4 dB. The employed energy-consumption model is based in that proposed by Feeney *et. al* [32]. This model registers the wireless card's state (transmit, receive, idle, or sleep), calculating the momentarily consumption according to the current state and the card's consumption associated to this state. In this case, the energy-consumption values presented in the BCM4330 chip's data-sheet [31] have been employed to characterize the simulated wireless cards (please, see Table I). It was assumed that the battery's output voltage is 3.6 V.

As aforementioned, the defined network has been evaluated in different node-motion conditions. Thus, nodes movement imitated the motion pedestrians by means of the Random Way Point model: each node in the network initiates each simulation at a random position and begins its motion toward a random point following a linear trajectory with a speed characterized by a Gaussian distribution with a mean value of 1.34 m/s and a standard deviation of 0.26 m/s [33]. Once the node reaches the destination point, it stops during a random time interval between 2 s and 5 s and, then, starts

moving to another different random point. Please observe that all nodes move during simulations, not only transmitters and receivers.

Regarding video transmission, the *UDPVideoStreamCli2* and the *UDPVideoStreamSvr2* modules included in the InetManet framework have been used. These UDP modules simulated a video-streaming communication in which the video server sends to the client UDP-traffic following a typical pattern for a video transmission. In the evaluated test-bench, a realistic 30 s video-trace available in [34] has been employed, corresponding to a video encoded at 30 frames per second (fps) with the H.264/SVC single layer VBR (Variable Bit Rate) codec. As the considered devices are cell-phones, the video frame size was 176 x 144 px (QCIF format), the GoP (Group of Pictures) 16, and the number of B frames between I and P frames is 1 (G16B1). Given these video features, the encoding system generated VBR frames with an average size of 485.5 Bytes, transmitting them at 118 Kbps. With this video coding configuration, it was intended to simulate a handset-oriented video-streaming application. As the 802.11 MAC layer's MTU (Maximum Transmission Unit) was set to its default value (2304 Bytes), it was not necessary to fragment each video frame into different packets. The starting time for each multimedia communication was generated randomly according to a Poisson distribution in a time interval from 0 to 10 s, after the convergence time for each routing protocol.

Focusing on quality measurement, QoS and QoE metrics have been taken into consideration. The latter have been calculated in terms of MOS (Mean Opinion Score), from the parametric model presented in [35] and the corrections to this model introduced in [36]. As shown in equation (1), this model allows obtaining accurate video-MOS estimations by means of the Packet Loss Rate (*PLR*), the coding scheme (*k*), the resolution (*a*) and the video-coding bit-rate (*Br*).

$$V_q = 1 + 4k \left(1 - \frac{1}{1 + \left(\frac{a \cdot Br}{v_1} \right)^{v_2}} \right) e^{-\frac{PLR}{v_3}} \quad (1)$$

In equation (1), *v*₁, *v*₂, and *v*₃ are experimental factors that depend on the coding scheme, the resolution applied, and the motion characteristics of the video. In this case, as described above, the widely extended H.264 codec has been used for the video-streaming transmission in QCIF format. It was assumed a video-content with low movement, as it is usual in video-telephony calls. Thus, the values assigned to these factors were: *k* = 1.12, *a* = 10.8, *v*₁ = 0.366, *v*₂ = 1.32, and *v*₃ = 3.5 respectively, hence consistent with the figures recommended by the authors [35], [36]. Regarding QoS metrics, in addition to the energy-consumption measurements pointed above, Packet Delivery Ratio (PDR) has been considered as well.

The remaining 802.11g parameters, such as DIFS (DCF Interframe Space), SIFS (Short Interframe Space), slot time, contention window (*CW*_{MIN}), and different-layers headers are shown in Table II, which additionally includes the main routing protocols time-outs.

Finally, the wireless transmission channel has been modeled as realistic as possible in order to avoid inaccuracies introduced by the simulation platform. Thus, the effect of fading channels on the wireless transmission medium has been considered by characterizing the physical layer with the

Table I
WIRELESS CARD ENERGY CONSUMPTION

State	Current Consumption
Transmit	250 mA
Receive	60 mA
Idle	1.25 mA
Sleep	0.18 mA

Table II
802.11g AND ROUTING PROTOCOLS PARAMETERS

Parameter	Bytes	Time
SIFS, DIFS, SLOT (μ s)	-	{10, 28, 9}
CWMIN (slots)	-	31
PLCP preamble (μ s)	-	4
{PLCP, MAC, SNAP} headers (μ s)	-, 28, 8	{16, 4.15, 1.18}
IP + UDP + RTP headers (μ s)	40	5.92
ACK (μ s)	14	2.07
BATMAN OGM interval (s)		1
BATMAN purge timeout (s)		200
OLSR {HELLO, TC, MID} intervals (s)		{2, 5, 5}
OLSR {Neighbor, Topology} hold times (s)		{6, 15}
AODV Active Route timeout (s)		3
AODV Hello interval (s)		1

well-known Nakagami- m model, which was shaped with an m value of 5 [37]. This characterization for the physical layer has been widely used in the related literature [38], [39] proving its accuracy. It has been also collected data when characterizing the physical layer with the Free Space propagation model. Thus, a comparison between the system behaviors in two different environmental hostility conditions has been performed.

IV. RESULTS

In this section, we provide a concise performance evaluation focused on the performance of the three routing protocols under study from both QoE and energy-consumption perspectives. Results include stationary and non-stationary scenarios. We also introduce a modification of the BATMAN protocol and its corresponding performance.

A. QoE vs. energy consumption study of BATMAN, OLSR, and AODV

Fig. 1 depicts the average QoE (in terms of MOS) obtained for different number of simultaneous video streams flowing through the afore-described static network managed by BATMAN, OLSR, or AODV. The average power consumed by each node during each entire simulation run (75 s) is also shown. Recall that these power-consumption values correspond only to wireless card consumption, ignoring other sources of battery draining. The results attained for two different levels of transmission-medium hostility are presented. Thus, Fig. 1(a) corresponds to the evaluation under ideal Free Space conditions, whereas Fig. 1(b) presents the results when characterizing the physical layer with the Nakagami- m propagation model (hostile conditions). Focusing on quality outcomes under different wireless channel conditions, the BATMAN protocol is the one with the most stable behavior. Observe that, although in the Free Space scenario (Fig. 1(a)) OLSR and AODV present the best results in terms of QoE (MOS always above 3.5), in the more realistic Nakagami- m scenario (Fig. 1(b)), the performance of these protocols dramatically drops, leading to QoE (MOS)

values below 1.5. These results can be explained from the information-dissemination strategy adopted by each protocol: Both OLSR and AODV make use of few control packets containing a big amount of information inside them; therefore, packet loss affecting these routing-data packets severely harms the respective path-finding algorithm. On the other hand, BATMAN uses a less controlled flooding algorithm [6]: based on the periodical (re) broadcasting of tiny OGM packets, the routing information gets finally distributed among all network nodes.

These strategies are also reflected in the energy consumed by each algorithm. Observe that, due to the continuous broadcasting and relaying of control packets, the BATMAN protocol is the one with the highest power consumption. In turn, OLSR has a more controlled routing-packets dissemination due to its characteristic Multi-Point Relaying (MPR) strategy [7]; thus, it presents considerably lower energy consumption than BATMAN. Finally, AODV is the protocol that employs less control packets because of its reactive nature. This is clearly noticeable in its power consumption, being the lowest one among the three protocols.

It is interesting to analyze the different behavior of each protocol's power-consumption pattern, depending on the hostility of the wireless transmission channel. While the proactive protocols (BATMAN and OLSR) consume more energy in the Free Space scenario, the reactive protocol (AODV) drains more energy in the hostile situation (please, compare Fig. 1(a) and Fig. 1(b)). This performance is explained as follows: BATMAN and OLSR make use of an

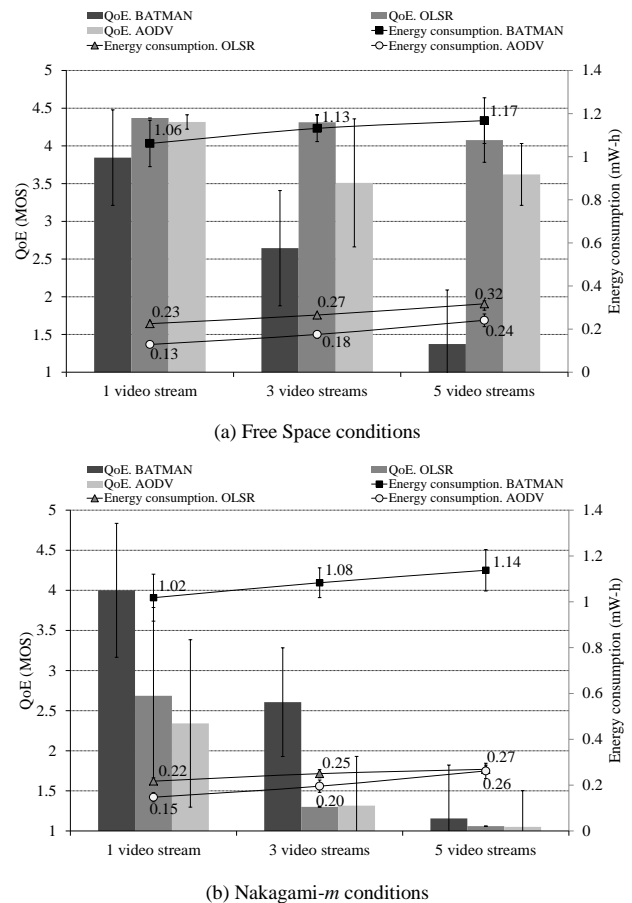


Fig. 1. QoE vs. Energy consumption in static scenario for a variable number of simultaneous video streams.

information-dissemination strategy based on the periodical (re) broadcasting of control packets. Thus, as in the Free Space environment few broadcast-packets are lost, a greater number of re-broadcastings occurs, increasing the energy consumption compared to the lossy Nakagami-*m* scenario. In turn, the AODV's route discovery mechanism makes nodes massively re-broadcast received route request (RREQ) packets until a route to the destination is established [40]. Thereby, in low-loss scenarios, i.e., Free Space, the routes are established quicker than in the hostile ones, so less control-packet broadcastings are necessary, hence reducing energy consumption.

In the lights of these results, it can be concluded that in environments with low level of packet loss, i.e., Free Space situations, there is not a clear trade-off between the quality attained by the multimedia communication and the needed energy for routing tasks. This statement is evidenced observing the QoE results attained by AODV, which are quite acceptable in terms of MOS compared with those obtained by the other routing protocols, while consuming low amount of energy. In lossy environments, our results suggest that the information-dissemination strategy of OLSR and AODV as a control-packet saving-strategy is clearly inefficient in this type of hostile scenarios, so other more energy-demanding routing approaches should be taken aiming at achieving higher levels of quality.

Regarding the protocols' capability to manage continuous route re-calculations, Fig. 2 shows the same evaluated parameters as before but under dynamic conditions. As described previously, all nodes follow a motion-pattern imitating the movement of a pedestrian handling a cell-phone [33]. Firstly, please observe the overall quality decrease affecting the three protocols (Fig. 2) compared to the static scenario (Fig. 1). Again, BATMAN is the protocol showing the greatest strength against hostile conditions. Both OLSR and AODV show much more degraded performance, particularly noticeable in the Free Space environment (Fig. 2(a)). In the Nakagami-*m* scenario (Fig. 2(b)), node motion additionally contributes to the already mentioned poor performance of these protocols under lossy conditions.

Focusing on energy consumption, there is a logical increase in the power consumed by the three protocols under study. Due to the continuous changes in the routes between transmitters and receivers, the routing tasks become more intense and hence the energy draining. The protocol's energy consumption behavior regarding the transmission-channel hostility is similar than that attained in the static scenario: proactive protocols, i.e., BATMAN and OLSR, drain more energy under Free Space conditions, whereas AODV (reactive protocol) consumes more in the lossy scenario.

B. Improving BATMAN performance with dynamic OGM intervals: Dyn-BATMAN

By default, BATMAN OGM interval is a fixed configurable time lapse, with a recommended value of 1 s following the standard guidelines [6]. However, authors discussed in a previous work about the need of properly tuning the BATMAN OGM interval in order to adapt the rate of control-packet broadcasts to current network conditions [9]. This idea has been already applied to other routing protocols under consideration in this work [41], [42]. Nevertheless, to the authors' knowledge, it has not been presented yet any dynamical adaption for the BATMAN

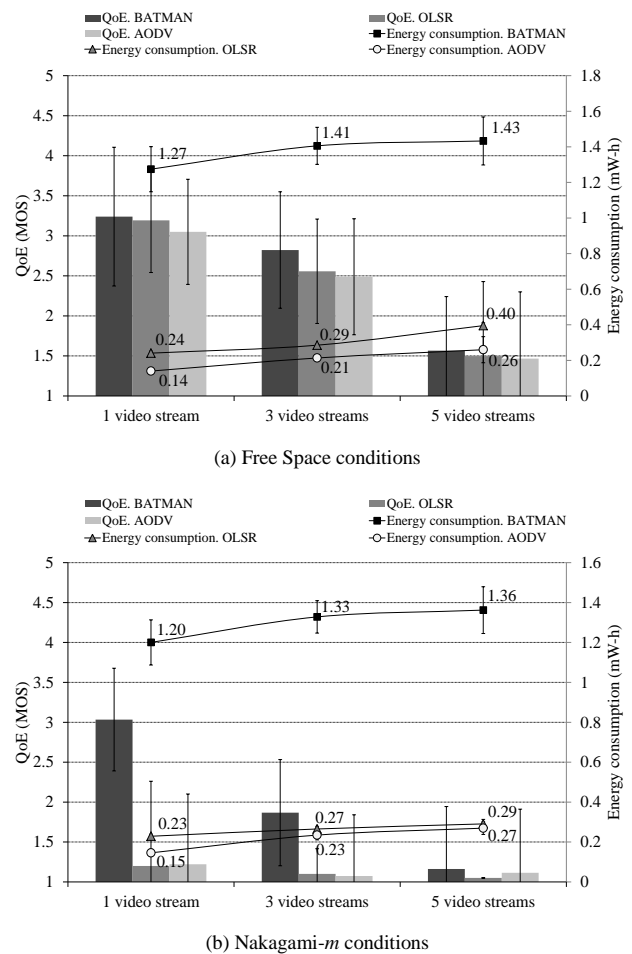


Fig. 2. QoE vs. Energy consumption in dynamic scenario for a variable number of simultaneous video streams.

routing protocol. In [9], it was demonstrated that an excessive number of OGM packets in the network noticeably harms the quality of the service flowing through it, especially in low-loss environments, e.g., Free Space. In that work, it was revealed that as the throughput in the network increased, it was necessary enlarging the time-interval between OGM broadcasts to maintain the desired quality level on the multimedia service crossing the network.

Consequently, we propose here to update the BATMAN Control-Message Sending Interval (CMSI) using equation (2) in order to control the number of OGM packets in the network. This new expression to update CMSI only employs the local information available in each node.

$$CMSI = a \cdot Throughput + b \tag{2}$$

In equation (2), *a* and *b* are experimental shaping factors and *Throughput* is the traffic crossing the node (in Kbps) since the last CMSI update. The CMSI is given in seconds. In order to obtain the values for the shaping factors a single-target experimental loop optimization strategy has been followed. To this end, extensive computer simulation has been carried out. Packet Delivery Ratio (PDR) has been selected as target function to maximize. After several preliminary experimental analyses, the considered range of values for factor *a* was [0.001, 0.008]; in turn, the selected values for the factor *b* were 1, 1.5, and 2. Fig. 3 shows the PDR attained in the scenarios under consideration when

applying the dynamic CMSI with the aforementioned ranges of shaping values. For the sake of simplicity, we only include here the results obtained with even values for factor a . For comparison purposes, the PDR values obtained by the original BATMAN have been also included (red line in Fig. 3). Thus, observe that in all evaluated scenarios, a proper tuning of the BATMAN CMSI leads to an increase in the average PDR in comparison with the original algorithm. Although the number of a and b factors combinations is high, there is one that stands out from the rest:

$$CMSI = 0.006 \cdot Throughput + 1.5. \quad (3)$$

This function has been marked in green in Fig. 3. By adjusting the dynamic CMSI's values to those showed in (3), the minimum time-interval between OGM's broadcasts is 1.5 s. This occurs when the throughput crossing the node is 0. Considering that the BATMAN's default OGM interval is 1 s, this enlargement of the minimum CMSI means a 33% reduction in the number of OGM packets in the network. Therefore, when applying the dynamic update of the CMSI, a controlled reduction of routing packets is achieved, resulting in an improvement on the performance of the service flowing through the network.

After selecting the most appropriate values for the CMSI, the performance of the dynamic BATMAN (Dyn-BATMAN) routing protocol needs to be compared with that of the original one. Thereby, Fig. 4 presents a comparison between the QoE (in terms of MOS) and the energy-consumption attained by the Dyn-BATMAN and the original BATMAN in static scenarios. Results for AODV have been also included as it previously showed the best trade-off between QoE and

energy consumption in some of the considered scenarios (please, see Fig. 1 and Fig. 2). Observe that in all simulated scenarios, the QoE (MOS) achieved by using Dyn-BATMAN is greater than that obtained by managing the network with the original BATMAN. In the scenario characterized by the Free Space propagation model (Fig. 4(a)), the superior performance of Dyn-BATMAN with 3 simultaneous video streams in the system is noticeable. Even so, under these ideal conditions, AODV maintains its superiority (in terms of MOS) in the most over-loaded scenario (5 simultaneous video streams). However, it is in the most realistic Nakagami- m scenario (Fig. 4(b)) where Dyn-BATMAN clearly presents the best performance compared with the other protocols; observe how Dyn-BATMAN shows its robustness to the adverse propagation conditions since it is the only protocol that reaches an acceptable QoE (MOS) value of 3 with 3 simultaneous video transmissions in the system. Under the heaviest traffic load, Dyn-BATMAN is the one that attains the highest QoE (MOS) level, although not sufficient to be considered as acceptable (QoE (MOS) below 3).

Regarding battery consumption, observe that in all simulated scenarios the power drained by Dyn-BATMAN is lower than that consumed by the original BATMAN, far from the lowest values achieved by AODV though. AODV is still the most battery-friendly protocol due to its reactive nature. The Dyn-BATMAN's power-consumption reduction is due to the decrease of OGM broadcasts by the proper enlargement of the CMSI. By applying this enhancement, wireless cards remain in an active state (transmitting or receiving) less time, entering in low-energy-consumption mode.

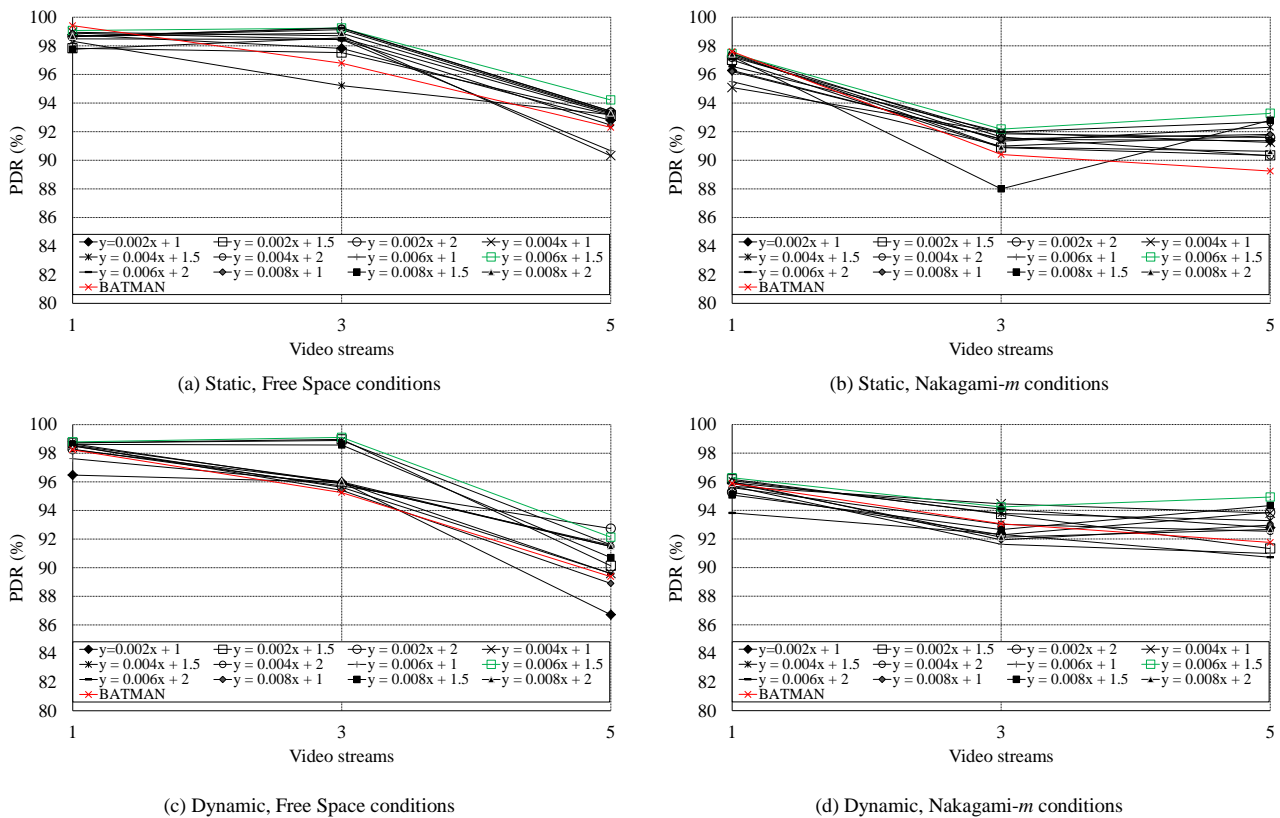


Fig. 3. Packet Delivery Ratio attained by the different adjustments applied to BATMAN CMSI in static ((a) and (b)) and dynamic ((c) and (d)) conditions for a variable number of simultaneous video streams. Free Space ((a) and (c)) and Nakagami- m ((b) and (d)) propagation models.

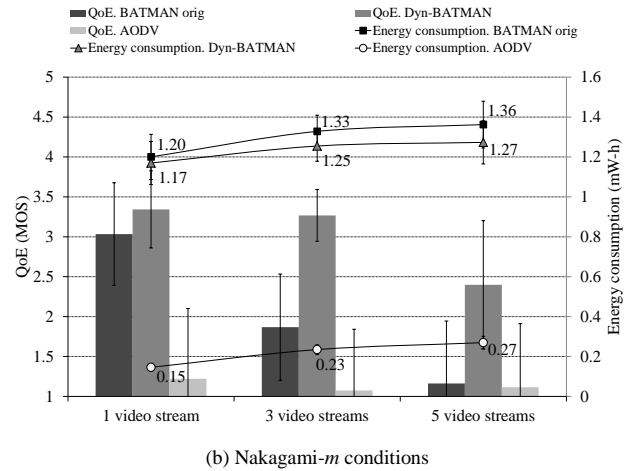
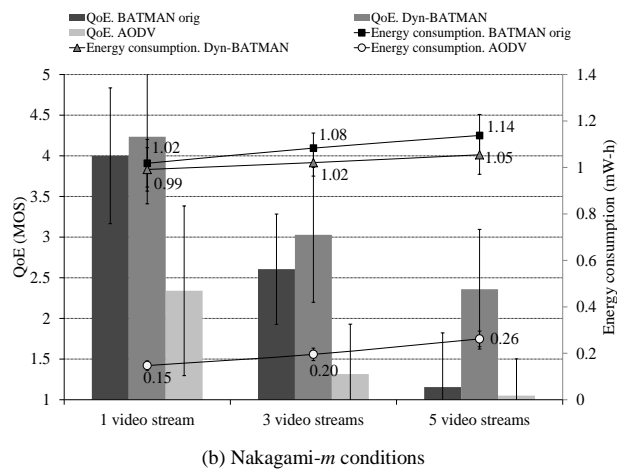
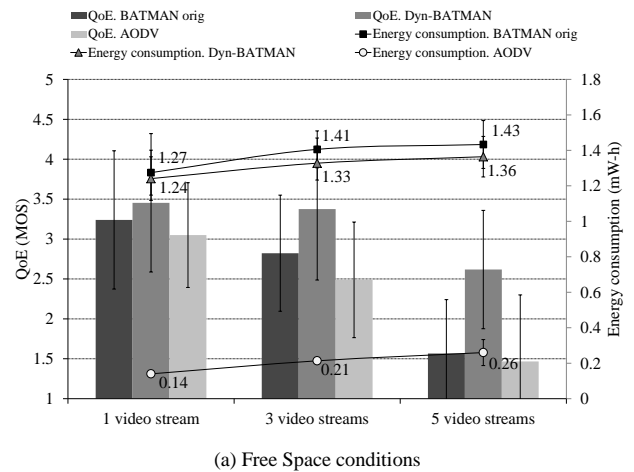
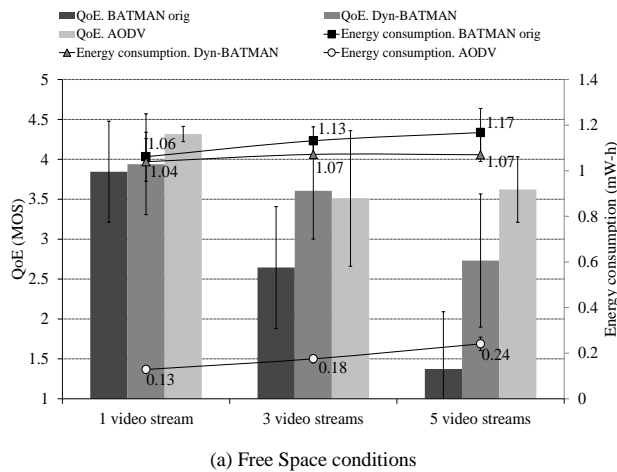


Fig. 4. Comparison of QoE and energy consumption among BATMAN orig, Dyn-BATMAN, and AODV in a static scenario with a variable number of simultaneous video streams.

Fig. 5. Comparison of QoE and energy consumption among BATMAN orig, Dyn-BATMAN, and AODV in a dynamic scenario with a variable number of simultaneous video streams.

This performance evaluation described for the static scenario (Fig. 4) is also valid for the dynamic one (Fig. 5). In fact, Dyn-BATMAN performs even better under node-motion conditions compared with the other 2 protocols. Note that the highest levels of QoE (MOS) are always achieved by managing the network with Dyn-BATMAN, no matter the number of simultaneous transmissions in the network or the wireless-channel conditions (Free Space, Fig. 5(a), or Nakagami- m , Fig. 5(b)). Although the QoE (MOS) values present a little decrease in comparison with the static scenario, the robustness showed against the difficulties posed by nodes motion is greater than that shown by the other protocols. Besides, the energy consumption remains under the levels achieved by the original BATMAN. Although the consumption of AODV continues in much lower figures than that obtained by Dyn-BATMAN, the quality level achieved by managing the network with AODV is now much lower than that attained by Dyn-BATMAN, too. Therefore, it can be concluded that the enhancement of Dyn-BATMAN in comparison with the original version of the protocol is twofold: (i) it permits an increase in the quality levels of the service crossing the network, together with (ii) a reduction of the energy consumed by the wireless cards. Thus, it has been demonstrated that a more efficient strategy of the ad-hoc routing protocols is possible, as it is feasible to obtain more battery-friendly routing algorithms without arming, even increasing, the quality level of the services in the network.

V. CONCLUSION

In this work, we have presented a comparison study evaluating the energy consumed by three prominent ad-hoc routing protocols (BATMAN, OLSR, and AODV) supporting a video-streaming service under different environmental conditions. Besides, the quality attained by the multimedia service in terms of QoE has been also considered. Thus, the trade-off between the quality perceived when consuming the video-service and the energy consumed by the end-device has been analyzed. From this study, it can be concluded that, on the one hand, in environments with low level of packet loss, i.e., Free Space situations, there is not a clear trade-off between the QoE attained by the multimedia communication and the needed energy for routing tasks. On the other hand, under lossy conditions, e.g., Nakagami- m environments, the best results in terms of QoE are achieved by the protocol that wastes more energy in their routing operations, i.e., BATMAN. Because of the robustness showed by BATMAN in adverse environmental conditions, an enhancement to this protocol has been proposed. This improvement consisted on the dynamic update of the Control-Message Sending Interval of OGM BATMAN messages (Dyn-BATMAN). With this new ability of dynamically adapting to current network conditions, BATMAN turns out a more battery-friendly protocol, even increasing the quality of the considered video service. Thus, it can be concluded that Dyn-BATMAN

presents an enhanced overall performance in comparison with the original version of the protocol. Dyn-BATMAN also overcomes OLSR and AODV under node-motion conditions. As future work, it is planned to conduct further tests to probe the validity of Dyn-BATMAN in other scenarios, as well as extending its capabilities with new enhancements that could permit increasing its performance supporting multimedia traffic.

ACKNOWLEDGMENT

This work was supported by the MINECO/FEDER project grant TEC2013-47016-C2-2-R (COINS).

REFERENCES

- [1] G. P. Perrucci, F. H. P. Fitzek, and J. Widmer, "Survey on energy consumption entities on the smartphone platform," in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–6.
- [2] S. Chiaravalloti, F. Idzikowski, and L. Budzisz, "Power consumption of WLAN network elements," *TKN Technical Report TKN-11-002. Tu Berlin*, 2011.
- [3] Cisco, "Cisco visual networking index: forecast and methodology, 2013–2018. White paper," *White Pap.*, 2014.
- [4] R. Sanchez-Iborra, M.-D. Cano, S. Moreno-Urrea, and J. J. P. C. Rodrigues, "QoE measurements and analysis for VoIP services," in *Handbook of Research on Networked Multimedia Communication Systems*, D. Kanellopoulos, Ed. IGI Global, 2015, p. In press.
- [5] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: a survey and tutorial on state of art, evolution and future directions of video quality analysis," *IEEE Commun. Surv. Tutorials*, vol. PP, no. 99, p. 1, 2014.
- [6] A. Neumann, C. Aichele, M. Lindner, and S. Wunderlich, "Better Approach To Mobile Ad-hoc Networking (B.A.T.M.A.N.)," *IETF Draft*, 2008.
- [7] T. Clausen and P. Jacquet, "Optimized Link State Routing (OLSR) Protocol," *RFC 3626*, 2003.
- [8] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," *RFC 3561*, 2003.
- [9] R. Sanchez-Iborra, M.-D. Cano, and J. Garcia-Haro, "Performance evaluation of BATMAN routing protocol for VoIP services: a QoE perspective," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 9, pp. 4947–4958, 2014.
- [10] "Information technology - MPEG Systems Technologies - Part 11: Energy-Efficient Media Consumption (Green Metadata), ISO/IEC JTC1/SC29/WG11 Final Draft International Standard 23001-11." 2014.
- [11] F. C. Fernandes, X. Ducloux, Z. Ma, E. Faramarzi, P. Gendron, and J. Wen, "The Green Metadata Standard for energy-efficient video consumption," *IEEE Multimed.*, vol. 22, no. 1, pp. 80–87, Jan. 2015.
- [12] C. Singhal, S. De, R. Trestian, and G.-M. Muntean, "Joint optimization of user-experience and energy-efficiency in wireless multimedia broadcast," *IEEE Trans. Mob. Comput.*, vol. 13, no. 7, pp. 1522–1535, Jul. 2014.
- [13] C. Singhal, S. De, R. Trestian, and G.-M. Muntean, "eWU-TV: user-centric energy-efficient digital TV broadcast over Wi-Fi networks," *IEEE Trans. Broadcast.*, vol. 61, no. 1, pp. 39–55, Mar. 2015.
- [14] S. Wang, W. Guo, C. Khirallah, D. Vukobratovic, and J. Thompson, "Interference allocation scheduler for green multimedia delivery," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2059–2070, Jun. 2014.
- [15] Y. Xu, R. Q. Hu, L. Wei, and G. Wu, "QoE-aware mobile association and resource allocation over wireless heterogeneous networks," in *IEEE Global Communications Conference*, 2014, pp. 4695–4701.
- [16] Z. Tang, Z. Wang, P. Li, S. Guo, X. Liao, and H. Jin, "An application layer protocol for energy-efficient bandwidth aggregation with Guaranteed Quality-of-Experience," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1538–1546, Jun. 2015.
- [17] E. Danish, V. Silva, A. Fernando, C. Alwis, and A. Kondo, "Content-aware resource allocation in OFDM systems for energy-efficient video transmission," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 320–328, Aug. 2014.
- [18] S. Khan, D. Schroeder, A. El Essaili, and E. Steinbach, "Energy-efficient and QoE-driven adaptive HTTP streaming over LTE," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 2354–2359.
- [19] "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," *ITU-T Recomm. J.144*, 2004.
- [20] Shivashankar, H. N. Suresh, G. Varaprasad, and G. Jayanthi, "Designing Energy Routing Protocol With Power Consumption Optimization in MANET," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 2, pp. 192–197, Jun. 2014.
- [21] S. Lim, C. Yu, and C. R. Das, "RandomCast: an energy-efficient communication scheme for mobile ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 8, no. 8, pp. 1039–1051, Aug. 2009.
- [22] F. De Rango, F. Guerriero, and P. Fazio, "Link-stability and energy aware routing protocol in distributed wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 4, pp. 713–726, Apr. 2012.
- [23] L. Chen and W. B. Heinzelman, "QoS-aware routing based on bandwidth estimation for mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 3, pp. 561–572, Mar. 2005.
- [24] T. Lu and J. Zhu, "Genetic algorithm for energy-efficient QoS multicast routing," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 31–34, Jan. 2013.
- [25] W. L. Huang and K. Ben Letaief, "Cross-layer scheduling and power control combined with adaptive modulation for wireless ad hoc networks," *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 728–739, Apr. 2007.
- [26] B. Tavli and W. B. Heinzelman, "Energy and spatial reuse efficient network-wide real-time data broadcasting in mobile ad hoc networks," *IEEE Trans. Mob. Comput.*, vol. 5, no. 10, pp. 1297–1312, Oct. 2006.
- [27] B. Tavli and W. B. Heinzelman, "Energy-efficient real-time multicast routing in mobile ad hoc networks," *IEEE Trans. Comput.*, vol. 60, no. 5, pp. 707–722, May 2011.
- [28] J.-L. Kuo, C.-H. Shih, C.-Y. Ho, and Y.-C. Chen, "A cross-layer approach for real-time multimedia streaming on wireless peer-to-peer ad hoc network," *Ad Hoc Networks*, vol. 11, no. 1, pp. 339–354, Jan. 2013.
- [29] Y. Peng, Y. Yu, L. Guo, D. Jiang, and Q. Gai, "An efficient joint channel assignment and QoS routing protocol for IEEE 802.11 multi-radio multi-channel wireless mesh networks," *J. Netw. Comput. Appl.*, vol. 36, no. 2, pp. 843–857, Mar. 2013.
- [30] "Omnnet++." [Online]. Available: <http://www.omnnetpp.org>.
- [31] Broadcom, "BCM4330 Preliminary Data Sheet," 2011.
- [32] L. M. Feeney and D. Willkomm, "Energy Framework: an extensible framework for simulating battery consumption in wireless networks," in *3rd International ICST Conference on Simulation Tools and Techniques*, 2012, p. 20.
- [33] D. Helbing, "Traffic and related self-driven many-particle systems," *Rev. Mod. Phys.*, vol. 73, no. 4, pp. 1067–1141, Dec. 2001.
- [34] "Video Trace Library." [Online]. Available: <http://trace.eas.asu.edu>.
- [35] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *IEEE International Conference on Communications*, 2008, pp. 110–114.
- [36] J. Joskowicz and J. C. L. Ardao, "Enhancements to the opinion model for video-telephony applications," in *5th International Latin American Networking Conference- LANC '09*, 2009, p. 87.
- [37] C. Oestges, N. Czink, B. Bandemer, P. Castiglione, F. Kaltenberger, and A. J. Paulraj, "Experimental characterization and modeling of outdoor-to-indoor and indoor-to-indoor distributed channels," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2253–2265, Jun. 2010.
- [38] V. Asghari, D. B. da Costa, and S. Aissa, "Performance analysis for multihop relaying channels with Nakagami-m fading: ergodic capacity upper-bounds and outage probability," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2761–2767, Oct. 2012.
- [39] D. Torrieri and M. C. Valenti, "The outage probability of a finite ad hoc network in Nakagami fading," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3509–3518, Nov. 2012.
- [40] M. Bouhorma, A. Boudhir, M. Ben Ahmed, and S. El Brak, "New route request mechanism for energy optimization in mobile ad hoc networks," in *19th Telecommunications Forum (TELFOR)*, 2011, pp. 230–233.
- [41] J. Garcia-Nieto and E. Alba, "Automatic parameter tuning with metaheuristics of the AODV routing protocol for vehicular ad-hoc networks," *Appl. Evol. Comput.*, vol. 6025, pp. 21–30, 2010.
- [42] E. Larsen, J. Flathagen, V. Pham, and L. LandMark, "iOLSR: OLSR for WSNs using dynamically adaptive intervals," in *Fifth International Conference on Sensor Technologies and Applications (SENSORCOMM 2011)*, 2011, pp. 18–23.

Testing Infrastructure for WebRTC applications

Boni García¹, Luis López-Fernández¹, Micael Gallego¹, Francisco Gortázar¹, Francisco Javier López²

¹Universidad Rey Juan Carlos. Camino del Molino S/N, 28943 Fuenlabrada, Spain
boni.garcia@urjc.es, luis.lopez@urjc.es, micael.gallego@urjc.es, francisco.gortazar@urjc.es

²Naevatec Inc. Calle Chile 10, 28290 Madrid, Spain
fjlopez@naevatec.com

Abstract- WebRTC is a new standard and industry effort that extends the web browsing model to directly exchange real-time media with other browsers. Despite the fact that WebRTC is an emergent technology and it is not completely mature at writing time, it is gaining the attention of practitioners quickly. For that reason, the mechanisms to provide quality assurance for WebRTC-based applications are key to release these kind of applications to production environment. This paper provides the description of the testing infrastructure developed within the open source project kurento.org. Kurento is a WebRTC media server and a set of APIs aimed to simplify the development of advanced video applications for the Web and smartphone platforms. Among these APIs, Kurento provides a high level testing infrastructure named Kurento Testing Framework (KTF). This framework simplifies the testing activities for any WebRTC application. It provides advanced features to perform complete assessment of WebRTC applications in terms of functionality, performance, and quality-of-experience.

Keywords: WebRTC, Software Testing, Performance, Quality-of-Experience.

I. INTRODUCTION

Web Real-Time Communications (WebRTC) is the umbrella term for several emergent technologies and APIs that aim to bring such communications to the Web. The standardization activity for WebRTC is split between the World Wide Web Consortium¹ (W3C) and the Internet Engineering Task Force² (IETF). On the one hand, W3C is defining the JavaScript APIs (Application Programming Interfaces) and the standard HTML5 tags to enable real-time media capabilities to browsers. On the other hand, IETF is defining the underlying communication protocols (SRTP, SDP, ICE, and so on) for the setup and management of a reliable communication channel between browsers [1].

Although still in its infancy, WebRTC is a technological initiative getting considerable worldwide attention. With the increasing development of WebRTC systems and services, the need for evaluating their quality becomes urgent.

WebRTC-based applications (or simply WebRTC applications) enable human-to-human communication. Therefore, these systems can be evaluated with respect to their multimedia conversation quality, i.e. whether they enable a good communication of information between the involved peers. In this line, this piece of research presents a

testing framework aimed to simplify the assessment of WebRTC applications.

This paper presents a testing framework aimed to simplify the testing process of WebRTC applications. This framework has been developed within the open source project Kurento [2], and so it is named Kurento Testing Framework (KTF).

The remainder of this paper is structured as follows. Section II presents the state of the art in two main topics of this research line: testing of web applications and quality-of-experience for multimedia. Section III summarizes some approaches for testing WebRTC available in the literature. Section IV introduces the WebRTC infrastructure (i.e. media server), since this work has been developed within the Kurento project (open source WebRTC media server). Section V details the testing framework for WebRTC created for Kurento, which can be used in general for any WebRTC application. Section VI provides implementation details and examples of test cases and results obtained with the proposed testing framework. Finally the conclusions and the future work is given on section VII.

II. BACKGROUND

A. Web Testing

Verification and Validation (V&V) -also known as Software Quality Control- is the disciplined approach to assessing software products and services. Software testing is the most commonly performed activity within V&V. Given a piece of code, software testing (or simply testing) consists of observing a sample of executions (test cases), and giving a verdict over them [3].

Testing of web applications shares the same objectives of traditional application testing, i.e. to ensure quality and finding defects in the required functionality and services. Due to its heterogeneity, web applications present important challenges for their quality assurance and testing [4].

In order to perform a complete assessment procedure, it is required to evaluate web applications from functional and non-functional perspectives. Functional requirements are actions that a software product must do to be useful to users. These requirements arise from the work that stakeholders need to do. Non-functional requirements (also known as quality attributes) are properties that the product must have. In other words, the functional requirements define what the system should do while non-functional requirements define how the system should behave [5]. According to Di Lucca

¹ <http://www.w3.org/TR/webrtc/>

² <http://tools.ietf.org/wg/rwcweb/>

and Fasolino, the most important non-functional requirements are performance, load, stress, compatibility, accessibility, usability, and security [6].

B. Quality-of-Experience for Multimedia

There is no agreement of what “software quality” actually means. The quality definition has been improved in 2005 at ISO 9000 standard, stating that quality is the “degree to which a set of inherent characteristics [...] fulfils requirements” [7].

In parallel to the re-consideration of the term “quality”, the term “Quality of Experience” (QoE) has gained momentum, mainly with respect to media transmission systems and services. This term was born to counter-balance the term Quality of Service (QoS) with something which addresses the user’s perceptions and experiences, because those were considered to be more appropriate for designing systems and services with a high acceptance [8].

The most widely way to classify QoE metrics is based on subjective or objective methods. Subjective methods are conducted to obtain information on the quality of multimedia services using opinion scores, while objective methods are used to estimate the network performance using models that approximate the results of subjective quality evaluation [9].

Subjective QoE measurement is time consuming, and is not particularly applicable in a production environment. Instead of directly collecting quality information, objective methods can be used to estimate the mean opinion score (MOS score) [10]. MOS score was first standardized by the International Telecommunication Union (ITU). With this metric, the system quality is assessed subjectively from the mean opinion on a five-point scale: 1 = bad, 2 = poor, 3 = fair, 4 = good and 5 = excellent. The MOS is used typically in subjective tests of audio listening (see Table I).

Table I
MOS 5-POINT QUALITY SCALE

Quality of the speech	Score	Impairment
<i>Excellent</i>	5	Imperceptible
<i>Good</i>	4	Perceptible, but not annoying
<i>Fair</i>	3	Slightly annoying
<i>Poor</i>	2	Annoying
<i>Bad</i>	1	Very annoying

The objective quality measurement methods can be classified into the following five main categories (see Fig. 1) [11][12]:

- Media-layer models: the QoE is calculated using the speech and video signal. This model does not require any information about the system under testing (e.g. codec type, packet-loss rate).
- Parametric packet-layer models: these models predict the QoE from the packet header information and do not analyze media signals.
- Parametric planning models: these models use quality planning parameters for networks and terminals to predict the QoE.
- Bit-stream-layer models: in these models, encoded bit-stream information and packet-layer information are used to measure QoE.
- Hybrid models: these models are a combination of previously mentioned technologies.

The media-layer objective quality measurement methods are subdivided into three types:

- Full Reference (FR): the degraded signal is compared pixel by pixel with the original signal.
- No Reference (NR): stream analysis on receipt without comparing it to the original signal. Quality information is extracted from the degraded signal, as no reference is available. These methods are very small, making them very suitable for analysis in real time.
- Reduced Reference (RR): these methods actually build upon representative parameters (typically statistical values) that allow for the quantification of the change of quality between the original and the distorted version.

The FR and RR are classified into three categories:

- Traditional point-based metrics. For example peak signal-to-noise ratio (PSNR)
- Natural visual characteristics oriented metrics. For example structural similarity (SSIM), video quality metric (VQM)
- Perceptual oriented metrics. For example perceptual evaluation of speech quality (PESQ), perceptual evaluation of video quality (PEVQ).

PSNR [13] is the proportion between the maximum signal power and the corruption noise power. This metric has several limitations and it can only be used as a measure of quality of reconstruction of loss compression codecs (e.g. for image compression).

SSIM [14] is a metric used for measuring the similarity between two images. The image quality measurement is based on a distortion free image as a reference.

VQM [15] has been developed by the Institute for Telecommunication Science (ITS) to provide an objective measurement for perceived video quality.

PESQ [16][17] is a method for evaluating speech quality autonomously as the experience of a telephony system user. PESQ is superseded by a listening-only signal based model called Perceptual Objective Listening Quality Assessment (POLQA) [18].

PEVQ [19] provides MOS scores of the video quality for IPTV, streaming video, mobile TV and video telephony. This method analyzes the degraded video signal through the network for evaluating the degradation.

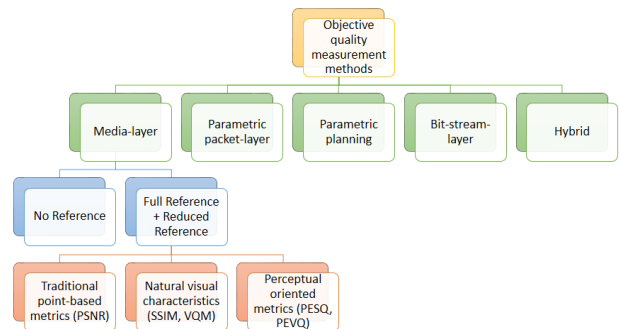


Fig. 1 Classification of objective measurement methods.

III. RELATED WORK

WebRTC is an emerging technology still in development. Regarding testing and quality of WebRTC applications, there not many references available in the literature yet.

For instance, Cinar and Melvin use a black-box testing technique to evaluate, via PESQ, the voice quality of WebRTC sessions under varying network delay and jitter [20]. In this paper, network emulators are employed to implement the delay and jitter variations. The results highlight the dangers of black-box testing, whereby test-bed issues can result in very misleading results.

Singh et al. evaluate the performance of the congestion control Receive-side Real-Time Congestion Control (RRCC), currently implemented in Google Chrome [21].

IV. WEBRTC SERVER INFRASTRUCTURE

WebRTC is a technology that provides Real-Time Communications capabilities to web browsers via JavaScript APIs. It has been conceived as a peer-to-peer architecture where browsers can directly communicate without the mediation of any kind of infrastructure.

This model is enough for creating basic applications but features such as group communications, media stream recording, media broadcasting or media transcoding are difficult to implement on top of it. For this reason, many applications require using a media server.

In the past few years, the expectations arising from WebRTC technologies have brought a golden era to media server vendors. The common features of WebRTC media servers fall into just three categories:

- Media bridging capabilities refer to attaining interoperability among networks or domains having incompatible media formats or protocols.
- Group communication capabilities include mixing and forwarding. This type of media server is called an MCU, following terminology from the ITU-T's H.323 recommendation.
- Media archiving capabilities deal with recording audiovisual streams into structured or unstructured repositories and recovering them later for visualization.

Kurento³ is a WebRTC media server and a set of client APIs making simple the development of advanced video applications for the Web and smartphone platforms. Kurento is a Free Open Source Software (FOSS) project (LGPL license). As a differential feature, the Kurento Media Server (KMS) provides advanced media processing capabilities involving computer vision, video indexing, augmented reality and speech analysis. Kurento modular architecture makes simple the integration of third party media processing algorithms (i.e. speech recognition, sentiment analysis, face recognition, etc.), which can be transparently used by application developers as the rest of Kurento built-in features.

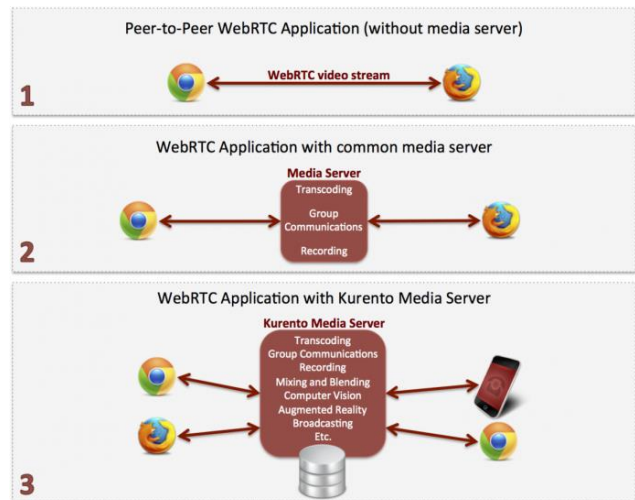


Fig. 2 KMS provides advanced media processing to the capabilities of common media servers.

Kurento also provides a complete testing framework (KTF, Kurento Testing Framework) aimed to simplify the assessment of WebRTC-based applications. KTF is not only for Kurento applications and can be used in general for WebRTC. Next section summarizes the main features of this testing framework.

V. WEBRTC TESTING FRAMEWORK

As introduced before, WebRTC enables web browsers with Real-Time Communications capabilities via JavaScript APIs. For this reason, in order to perform tests for WebRTC applications, it is a must to be able to automate test execution using real web browsers (Chrome, Firefox, Internet Explorer, and so on). The well-known open source testing framework Selenium⁴ is capable of handling this task. Selenium is composed by three components:

- Selenium IDE: It is a Firefox plugin implementing the record and playback pattern for web applications.
- Selenium WebDriver. It allows to handle local browsers natively as a user would using different programming languages (e.g. Java, C#, Python, Ruby, PHP, Perl, or JavaScript). It can be seen as the evolution of the now deprecated project Selenium RC (Remote Control).
- Selenium Grid: It allows distributing browser execution on parallel remote machines.

KTF provides a high level framework to perform complete automated testing for WebRTC-based applications. In the core of this framework, Selenium is a key component since it allows to handle browsers from the test logic. KTF supports three kind of browsers (scope):

- Local browsers. The host running tests should have installed web browsers in the operating system.
- Remote browsers. The execution of a test can be configured to run in a remote browser. These tests are implemented using Selenium Grid.

³ <https://www.kurento.org/>

⁴ <http://www.seleniumhq.org/>

- Remote browsers from SauceLabs⁵. SauceLabs is a PaaS (Platform as a Service) cloud solution to support remote testing based on Selenium. It provides a set of cloud instances able to run WebDriver tests. At writing time, SauceLabs supports 480 combinations of platform (e.g. Linux, Windows, Mac OS X, Android, iOS), browser (Chrome, Firefox, Internet Explorer, etc), and browser version.

A. Test Scenario

The first high-level concept introduced by KTF is the **test scenario**. Test scenario can be seen as the client-side infrastructure running a test case or suite. In other words, it is the set of browsers which are going to exercise the web application under test.

Let's suppose one of the simplest WebRTC application possible, in which two peers establish a real time communication through the Web. A textual description of a test scenario example would be as follows: "peer one uses Chrome and peer two uses Firefox". Another test scenario could be "both peers first use Chrome, and then use Firefox".

KTF allows to setup different tests scenarios based on a custom and highly customizable JSON notation. In these JSON files, several test executions can be setup. For each execution, the browser scope can be chosen, i.e. local browsers, remote browsers, or browsers provided by SauceLabs.

For example, listing I shows a test scenario in which two executions are defined. First execution defines two local browsers (identified as *peer1* and *peer2*), Chrome and Firefox respectively. The second execution defines also two browsers, but this time browsers are located in the cloud infrastructure provided by SauceLabs (Internet Explorer 11 on Windows 8.1 and Safari 8 Mac OS X Yosemite).

Listing I
EXAMPLE OF TEST SCENARIO IN JSON NOTATION

```
{
  "executions" : [
    {
      "peer1" : {
        "scope" : "local",
        "browser" : "chrome"
      },
      "peer2" : {
        "scope" : "local",
        "browser" : "firefox"
      }
    },
    {
      "peer1" : {
        "scope" : "saucelabs",
        "browser" : "explorer",
        "version" : "11",
        "platform" : "win8_1"
      },
      "peer2" : {
        "scope" : "saucelabs",
        "browser" : "safari",
        "version" : "36",

```

```
        "platform" : "yosemite"
      }
    ]
  }
}
```

Thus, KTF allows in simple fashion to setup rich test scenarios to perform WebRTC testing using real browsers. But this is not enough to perform a complete verification of a WebRTC application. Testing activities are aimed to provide answers about a System Under Test (SUT) to developers and testers [22]. Regarding WebRTC, many questions should be addressed before releasing an application to a production environment. For example: "Are WebRTC peers always receiving media?", "Is my application compatible with different versions of web browsers?" "What happens to my system when the number of WebRTC peers increase?", "Is the quality of the received media perceptible for users?".

In order to answer these questions (in other words, in order to perform different kind of tests), KTF provides further high-levels mechanisms to perform complex assessment activities easily.

B. Testing Activities

As introduced before, the aim of web applications testing consists of executing a web application under test using combinations of input in order to ensure quality and reveal failures. KTF provides specific capabilities to perform:

- Functional test. Assessment for WebRTC media capabilities.
- Performance tests. Evaluation of system behavior whilst web application is exercised with many concurrent requests.
- Quality-of-experience tests. These kinds of tests assess the quality of the media received in the browsers using QoE methods as depicted on section II.B.

Regarding **functional** tests, two out-of-the-box capabilities are provided by KTF. WebRTC media is played in browsers using the HTML5 video element. Several media events are triggered by the video tag, for instance, play, pause, seek, change of volume, mute, or change of the playback rate. KTF allows subscribing to any of these video tags events, asserting if the subscribed events happens in a given time. In addition, the playing time is also captured by KTF. Therefore tests can easily assess whether or not the playback time is as supposed.

Another functional feature provided by KTF is the capability of analyzing the color of the received media in HTML5 video tags. Given the Cartesian coordinates within the video tag, KTF is able to get the RGB (Red, Green, Blue) color components. Then, a built-in comparison method based on the calculation of the RGB distance from the expected color and the real color can be used to assess whether or not the received color is similar to the expected one. In other words, the distance (*d*) calculated using equation I should be less than a given threshold (*th*).

$$(I) \quad d = \sqrt{(R_{real} - R_{expected})^2 + (G_{real} - G_{expected})^2 + (B_{real} - B_{expected})^2} < th$$

KTF provides a fluent API to use these testing capabilities. See section VII for further details and examples.

⁵ <https://saucelabs.com/>

Regarding **performance** tests, KTF provides two main capabilities. First, a built-in function to handle the rate of the browser interactions with the SUT in form of a ramp (see Fig. 3). Given a test scenario composed by N browsers, each browser interacts with the SUT one by one at a rate given by a customizable parameter (*parallel.browsers.rate*). The simultaneous execution of all browsers is maintained a given number of seconds given by another parameter (*parallel.browsers.holdtime*).

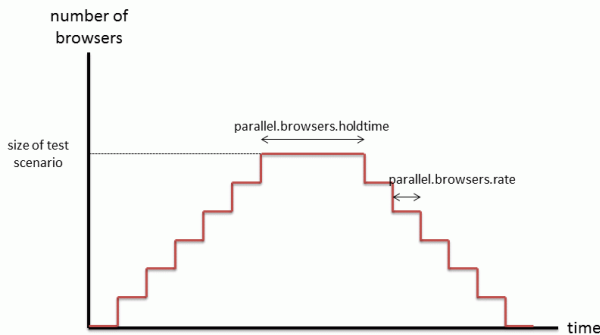


Fig. 3 Browser ramp.

Let's see an example of a JSON test scenario for this kind of tests. A typical WebRTC application could be a video room application in which a peer (called *presenter*) sends its media stream to N peers (called *viewers*). In order to test this kind of application using the performance ramp described before, we first need to define the test scenario. Listing II shows how to define the test scenario for this video room application (with 80 remote viewers).

Listing II
EXAMPLE OF SCENARIO FOR PERFORMANCE TESTS

```
{
  "executions" : [
    {
      "presenter" : {
        "viewer" : "local",
        "browser" : "chrome"
      },
      "viewer" : {
        "scope" : "remote",
        "browser" : "firefox",
        "instances" : 80
      }
    }
  ]
}
```

Moreover, KTF provides the capability of monitoring a given machine, typically the sever hosting the SUT. The parameters gathered by the monitor are the following:

- Time (relative to the start of the test).
- Number of incoming clients.
- CPU usage (percentage).
- Memory usage (number of bytes and percentage out of the total).
- Swap memory usage (number of bytes and percentage out of the total).

- Network interfaces usage (number of sent and received bytes in each of the network interfaces)

This information can be gathered locally, i.e. the SUT is running in the same host that it is executing the performance test or remotely, i.e. the SUT is running in a different host.

In addition, the system monitor also gathers client-side latency measurement. KTF implements a novel procedure to calculate the end to end latency in the WebRTC media transmission. This method is based on color detection of the local (sent by a peer) and remote (received by another peer).

To implement this method, first of all we have to detect the color change in the local and remote stream. Fig. 4 shows an example of a video⁶ with the required features. This video has a duration of 15 seconds. The first 5 seconds in the upper-left part a completely red video is shown. After that, another 5 seconds green video is shown. Finally another 5 seconds blue video is sent. This video is sent in loopback with the source peer in the WebRTC communication. The peer(s) consuming that media receives this video with a given latency.

As depicted before, KTF is able to recognize colors in the WebRTC streams. Thus, the color change is detected simultaneously in the local and remote stream. After a synchronization step, KTF is able to calculate the network latency each 5 seconds simply calculating the color change time from the remote and local stream.

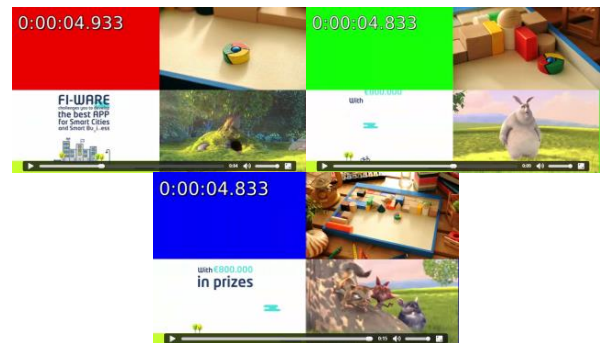


Fig. 4 Example of video test used to measure latencies.

Regarding **quality** tests, currently PESQ (Perceptual Evaluation of Speech Quality) is supported to evaluate the received audio quality. To achieve this, first the PESQ algorithms needs to compare an original audio source to a degraded version.

For this reason, in the JSON test scenario description, a new field called *audio* can be added. This parameter points to a WAV file that will be merged in the WebRTC peer acting as source of the communication by means of the Web Audio API⁷. In the WebRTC peer consuming the media, the audio is recorded by means of FFmpeg⁸. Finally, the quality test infrastructure executes PESQ with the source and the received audio file. The result is a MOS score from 1 to 5 as described on Table I. This figure can be used to assert when a

⁶ <http://files.kurento.org/>

⁷ <https://dvcs.w3.org/hg/audio/raw-file/tip/webaudio/specification.html>

⁸ <https://www.ffmpeg.org/>

test fails when the resulting MOS is lower than a given results (e.g. when resulting MOS is 3 or less).

VI. USING THE KURENTO TESTING FRAMEWORK

KTF has been implemented in Java. Moreover, KTF has been designed to use JUnit framework to implement tests. KTF dependency can be imported using Maven by means of the project coordinates described on listing III.

Listing III
KURENTO TESTING FRAMEWORK MAVEN DEPENDENCY

```
<dependency>
  <groupId>org.kurento</groupId>
  <artifactId>kurento-test</artifactId>
  <version>6.0.0</version>
  <scope>test</scope>
</dependency>
```

In order to create new JUnit tests, these tests should extend the parent test class `KurentoTest` (located in the package `org.kurento.test.base.KurentoTest`). Listing IV shows an example of JUnit Functional test. This listing provides important details to understand how tests which uses KTF works. These tests should use the parameterized feature of JUnit. This parameter accepts the test scenario (see constructor), which is created from the JSON file `browsers.json`. This JSON file follows the notation depicted in section V.A. and should be available in the project classpath.

Listing IV
JUNIT TEST EXAMPLE

```
import org.junit.Test;
import org.junit.runners.Parameterized.Parameters;
import org.kurento.test.base.KurentoTest;
import org.kurento.test.config.TestScenario;

public class MyTest extends KurentoTest {

    public MyTest(TestScenario testScenario) {
        super(testScenario);
    }

    @Parameters(name = "{index}: {0}")
    public static Collection<Object[]> data() {
        return TestScenario.json("browsers.json");
    }

    @Test
    public void test() {
        // Test logic
    }
}
```

A. Functional Tests

Let's suppose the JSON file with the test scenario description is the same as the depicted on listing I. In this case, in the test logic implementation, the browsers can be accessed by means of its identifiers, i.e. `peer1` and `peer2`, as illustrated in listing VI. This snippet shows an example of subscription and assertion to the `playing` media event in the HTML5 video tag with id `video` in browser `peer1`. Then, for the same video tag, the media in the position `x=0, y=0` is read and another assertion is made. Finally the current time for that video tag is also read.

Listing V
KTF FUNCTIONAL SNIPPETS

```
// Media events
getBrowser("peer1").getVideoTag("video").
  subscribeEvents("playing");
boolean playing = getBrowser("peer1").
  getVideoTag("video").waitForEvent("playing");
Assert.assertTrue(playing);

// Color
Color realColor =
getBrowser("peer1").getVideoTag("video").
  getColorAt(0,0);
Assert.assertTrue(similarColor(realColor,
  expectedColor));

// Time
double currentTime = getBrowser("peer1").
  getVideoTag("video").getTime();
Assert.assertTrue(compareTime(currentTime,
  expectedTime));
```

B. Performance Tests

KTF provides an API to perform the browser ramp as depicted in Fig. 3. Listing VII shows an example of this method, used in a test scenario as defined in listing II (1 presenter and 80 viewers).

Listing VI
USAGE OF KTF API TO CARRY OUT PERFORMANCE TESTING

```
private SystemMonitorManager monitor;

@Before
public void setup() {
    String host = "127.0.0.1";
    String login = "user";
    String key = "/path/to/key.pem";
    monitor = new SystemMonitorManager(host,
        login, key);
    monitor.start();
}

@After
public void teardown() {
    monitor.stop();
    monitor.writeResults("results.csv");
    monitor.destroy();
}

@Test
public void test() {
    Map<String, BrowserClient> browsers =
        getTestScenario().getBrowserMap("viewer");

    // Test logic for presenter

    ParallelBrowsers.ramp(browsers, monitor, new
        BrowserRunner() {
            public void run(BrowserClient browser)
                throws Exception {
                // Test logic for viewers
            }
        });
}
```

The class `SystemMonitorManager` performs the analysis on a host (typically the machine hosting the SUT).

The results is tabulated in a CSV file, useful to get charts of the system performance with respect to the time or number of concurrent users. Figures 5 to 8 shows an example of charts obtained from a performance test execution in ramp.

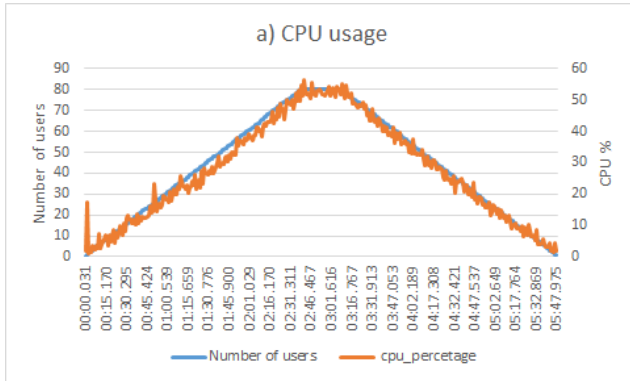


Fig. 5 CPU usage vs. number of user.

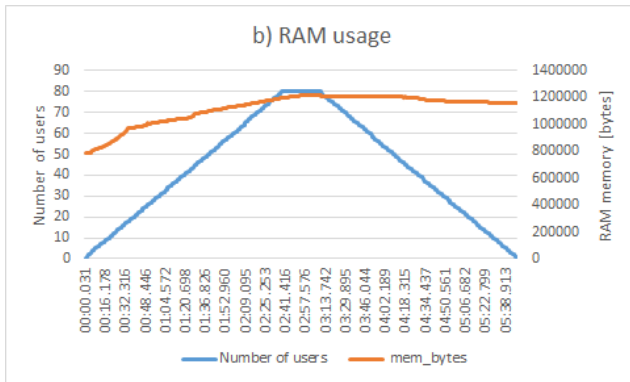


Fig. 6 Memory usage vs. number of users.

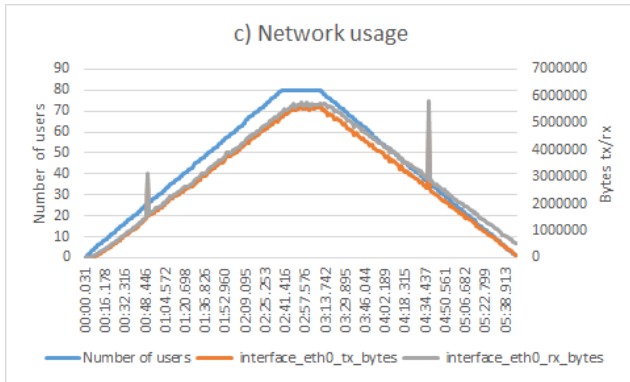


Fig. 7 Network usage vs. number of users.

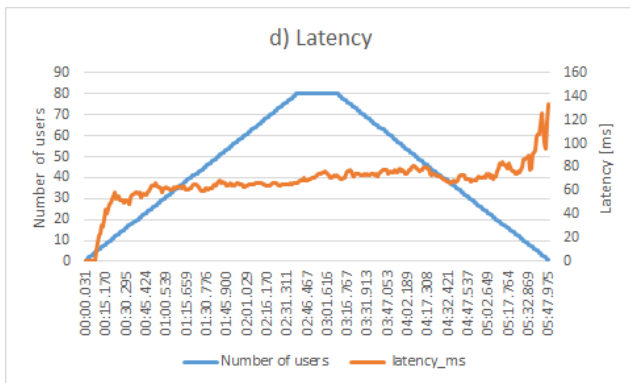


Fig. 8 Network latency vs. number of users.

Figure 5 shows how the number of users (left Y-axis) together with the percentage of CPU (right Y-axis) vary depending on the time (X-axis). With this chart we can check than the relationship between these two variables (number of users and CPU usage) is straight.

Figure 6 shows how the number of users (left Y-axis) together with the number of bytes of RAM memory (right Y-axis) vary depending on the time (X-axis). With this chart we see that RAM memory grows slowly as the number of users is increased.

Figure 7 shows how the number of users (left Y-axis) together with the number of bytes sent and received (right Y-axis) vary depending on the time (X-axis). Except for some little peak in the reception network interface, it can be seen that the relationship among these three variables (number of users, number of bytes sent, and number of bytes received) is directly related.

Finally figure 8 shows how the number of users (left Y-axis) together with the end-to-end latency in milliseconds (right Y-axis) vary depending on the time (X-axis). This result is quite interesting since the latency remains quite constant (around 70 milliseconds in mean) as the number of users first grow and then decrease.

All in all, in the light of these examples we conclude that the system performance (in terms of CPU/RAM usage, bytes sent/received and latency) is acceptable for a number of 80 concurrent users.

C. Quality Tests

Snippet shown in listing VIII shows the usage of the API provided by KTC to carry out QoE evaluation based on PESQ. When a custom audio file is specified in the test scenario, automatically KTF records the audio received by WebRTC. Then, using the class Recorder, the test logic can compare the recording with the original source by means of an integrated implementation of the PESQ algorithm. As a result, a MOS score is obtained, which can be used to feed a test assertion.

Listing VII
USAGE OF KTF API TO PERFORM QUALITY TESTING

```
int sampleRate = 16000; // samples per second
float minPesqMos = 3; // PESQ MOS [1..5]
String audioUrl =
"http://files.kurento.org/audio/10sec/fiware_mono_16khz.wav";

float realPesqMos = Recorder.getPesqMos(audioUrl,
sampleRate);
Assert.assertTrue(realPesqMos >= minPesqMos);
```

VII. CONCLUSIONS AND FUTURE WORK

Testing of WebRTC applications presents important challenges for practitioners. Due to the human-to-human nature of the WebRTC applications, the QoE measurement provides a key tool in order to evaluate the perceived quality in WebRTC applications.

This piece of research presents a high-level testing framework to perform complete assessment of WebRTC-based applications. This framework has been created within the open source project Kurento, and so it is called Kurento Testing Framework (KTF).

KTF provides several mechanisms to evaluate functional parameters (media events, color detection), performance (system monitor, latency measurement based on the color comparison of media sent and received), and quality-of-experience (evaluation of the audio quality by means of PESQ).

KTF uses Selenium WebDriver/Grid to perform automated interaction with WebRTC applications. It handles a custom JSON notation to specify the test scenario. This scenario describes how many and where it is going to be located the browsers used to perform the tests. The possibilities are: local and remote. This JSON allows to perform compatibility tests easily, simply specific different browsers in different platforms. In addition is quite useful to setup different test executions in continuous integration environments.

Kurento is an active project, and so is the Kurento Testing Framework. New capabilities are planned to be implemented in KTF in the short term. First, some objective quality measurement methods are going to be integrated as video QoE assessment (e.g. PEVQ, VQM, or SSIM). In addition, WebRTC statistics⁹ are going to be gathered by the system monitor in order to evaluate the characteristics of the media exchanged in the WebRTC communication.

ACKNOWLEDGEMENTS

This work has been supported by the European Commission under projects FI-WARE FP7-2011-ICT-FI, GA-285248, and NUBOMEDIA FP7-ICT-2013-1.6, GA-610576.

REFERENCES

- [1] S. Loreto, S. P. Romano, "Real-Time Communication with WebRTC", O'Reilly Media, 2014.
- [2] L. López-Fernández, M. París, R. Benítez, F.J. López, J. Santos, "Kurento: a media server technology for convergent WWW/mobile real-time multimedia communications supporting WebRTC", 14th IEEE International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1-6, 2013.
- [3] A. Bertolino, "Software Testing Research: Achievements, Challenges, Dreams", in 'Future of Software Engineering' (FOSE '07), pp. 85-103, 2007.
- [4] Y.F. Li, P.K. Das, D.L. Dowe, "Two decades of Web application testing - A survey of recent advances", Information Systems, Volume 43, pp. 20-54, 2014.
- [5] S. Robertson, J. Robertson, "Mastering the Requirements Process", 2nd Edition, Addison-Wesley Professional, 2006.
- [6] G.A. Di Lucca, A.R. Fasolino, "Testing web-based applications: The state of the art and future trends". Information Software Technology no. 48(12), pp. 1172-1186, 2006.
- [7] International Organization for Standardization, "ENISO 9000: Quality management systems - fundamentals and vocabulary", Geneva, 2005.
- [8] S. Möller, A. Raake, "Quality of Experience. Advanced Concepts, Applications and Methods", Springer, 2014.
- [9] A. Mellouk, S. Hoceini, H.A. Tran, "Quality of Experience for Multimedia. Application to Content Delivery Network Architecture", Wiley, 2013.
- [10] ITU Telecommunication, "ITU Recommendation P.801: Mean Opinion Score (MOS) terminology", 2006.
- [11] S. Chikkerur, V. Sundaram, M. Reisslein, et al., "Objective video quality assessment methods: a classification, review, and performance comparison", IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 165-182, 2011.
- [12] A. Takahashi, D. Hands, V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV", IEEE Communications Magazine, vol. 46, no. 2, pp. 78-84, 2008.
- [13] Q. Huynh-Thu, M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment", Electronics Letters, vol. 44, no. 13, pp. 800-801, 2008.
- [14] Z. Wang, A. Bovik, H. Sheikh, et al., "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.
- [15] M. Pinson, S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality", IEEE Transactions on Broadcasting, vol. 50, no. 3, pp. 312-322, 2004.
- [16] A. Rix, M. Hollier, A. Hekstra, et al., "Perceptual evaluation of speech quality (PESQ): the new ITU standard for end-to-end speech quality assessment part I-time-delay compensation", Journal of the Audio Engineering Society, vol. 50, no. 10, pp. 755-764, 2002.
- [17] ITU Telecommunication, "ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", 2001.
- [18] ITU Telecommunication, "ITU-T Recommendation P.863: Perceptual objective listening quality assessment", 2001.
- [19] ITU Telecommunication, "ITU-T Recommendation J.247: Objective perceptual multimedia video quality measurement in the presence of a full reference", ITU Telecom, 2008.
- [20] C. Yusuf, M. Hugh, "WebRTC quality assessment: Dangers of black-box testing," Digital Technologies (DT) 10th International Conference, vol. 32, no. 35, pp. 9-11, 2014.
- [21] V. Singh, A.A. Lozano, J. Ott, "Performance Analysis of Receive-Side Real-Time Congestion Control for WebRTC", Packet Video Workshop (PV) 20th International, vol. 1, no. 8, pp. 12-13, 2013.
- [22] G.M. Winberg, "Perfect software and other illusions about testing", Dorset House Publishing Co Inc., 2008.

⁹ <http://www.w3.org/TR/webrtc-stats/>

Maqueta para la evaluación de la calidad de experiencia del servicio videostreaming 3D en LTE

Inmaculada Palomo, Marta Solera Delgado, Matías Toril, Javier Poncela-González, Gerardo Gómez.

{inmapalomo,msolera,mtoril,javier.gomez}@ic.uma.es

Dpto. de Ingeniería de Comunicaciones. ETSI Telecomunicación. Universidad de Málaga.

Campus de Teatinos s/n 29071 Málaga (España)

Abstract-With the deployment of the first commercial LTE (*Long Term Evolution*) networks, mobile operators need to understand how QoS (*Quality of Service*) network and codec parameters affect the subjective quality of high-definition and 3D video, as perceived by customers. In this article, we describe the development of a testbed system for evaluating the QoE (*Quality of Experience*) of 3D Video Streaming Service over LTE. The proposed system consists of three elements: the streaming server, the IP mobile network emulator and the streaming client. In the testbed, different network conditions are configured by setting network emulator parameters based on the results obtained by a system-level LTE simulator. Results show how average network load and user position inside the cell can have a strong impact on the QoS and QoE perceived by the end 3D video user.

I. INTRODUCCIÓN

En los últimos años, la industria de la electrónica de consumo ha dedicado numerosos esfuerzos a llevar el video y la televisión tridimensional (3D) a los hogares. En paralelo, las industrias del cine y del deporte también han apostado por las tres dimensiones con la creación de contenidos 3D atractivos para el espectador. Así, un estudio de CISCO, principal fabricante de dispositivos de red, prevé que en 2014 el tráfico de vídeo de alta definición y 3D alcance el 46% del tráfico total en Internet.

Lamentablemente, las redes móviles tradicionales no están diseñadas para la distribución de video de alta definición y 3D. Para que estos formatos se consoliden, es preciso distribuir de forma eficiente la gran cantidad de información que requieren ambos formatos. Por ello, se hacen necesarias nuevas formas de codificar, comprimir y transportar este tipo de datos.

Para evaluar el rendimiento de sus redes, los operadores de telefonía móvil consideraban hasta hace poco medidas objetivas de calidad de servicio (*Quality of Service*, QoS), como el retardo o la velocidad de transmisión. No obstante, aunque estas medidas están relacionadas con el grado de satisfacción del usuario, la relación exacta no es directa. Por ello, los operadores están cambiando la estrategia clásica de gestionar la red, centrada en el control del rendimiento de la red, por una estrategia que ponga el foco en satisfacer al usuario.

En este contexto, el reciente despliegue de las redes móviles comerciales de cuarta generación (4G), basadas en el estándar LTE (*Long Term Evolution*), requiere la realización de estudios que determinen el impacto real de los parámetros de QoS y de la compresión en la experiencia del servicio de distribución de vídeo 3D en este tipo de redes.

En trabajos anteriores, se ha estudiado la calidad de experiencia en la distribución de vídeo 3D sobre redes IP cableadas [1][2] e inalámbricas [3]. Otros trabajos han estudiado la distribución de vídeo convencional 2D en redes móviles [4]. Los autores, en un artículo previo [5], presentan una plataforma de sistema de distribución de vídeo 3D en tiempo real sobre IP. Con dicha maqueta, se puede evaluar con usuarios reales el impacto de los parámetros de calidad de servicio definidos en la red (p.ej., retardo, tasa de pérdidas) y los parámetros configurables a nivel de aplicación (p.ej., codificador, tipo de contenido) sobre la calidad de experiencia (*Quality of Experience*, QoE) percibida por el observador. Su principal novedad frente a maquetas similares descritas en la bibliografía (p.ej., [6][7]) es la combinación del tráfico de vídeo 3D con la emulación de una red celular. En ese artículo se presentaba un ejemplo sencillo de configuración de la maqueta para evaluar la calidad del servicio de *videostreaming* 3D en una red LTE.

Este trabajo complementa al anterior, profundizando en la configuración de los elementos que conforman la maqueta. Por un lado, se presentan nuevas modificaciones en el emulador de red para modelar mejor las características del retardo de paquetes cuando el proceso de llegada de paquetes es discontinuo en el tiempo, como es el caso del servicio de *videostreaming*. Por otro lado, se describe una nueva metodología de ajuste de los parámetros del emulador de red basándose en los resultados obtenidos con un simulador de nivel de sistema. Una vez configurado el emulador, se evalúan 4 escenarios diferentes que pretenden reflejar la experiencia de un usuario del servicio de *videostreaming* 3D cuando se sitúa en el centro o en el borde de la celda, en condiciones de baja o alta carga de red.

El resto del artículo se estructura de la siguiente manera. En la sección II se hace una breve introducción a la maqueta de emulación de red LTE para el servicio de *videostreaming*, describiendo la configuración de las herramientas utilizadas. La sección III explica la simulación de la interfaz radio LTE. La sección IV cuenta la metodología experimental de prueba del sistema. La sección V muestra los resultados y, por último, la sección VI expone las principales conclusiones.

II. SISTEMA DE EMULACIÓN DE VIDEOSTREAMING EN LTE

Existen diferentes metodologías para evaluar el comportamiento extremo a extremo de una red de telecomunicación. La primera de ellas es la simulación, que utiliza un modelo simplificado de un escenario real y que proporciona unos resultados limitados a la complejidad del modelo. Una segunda alternativa es la emulación, que, dados

unos indicadores de rendimiento generales de la red, imita su comportamiento en un entorno controlado. La tercera alternativa es la realización de medidas de campo, que se toman directamente de los elementos de la red en un escenario real. La plataforma descrita en este trabajo combina todas estas estrategias, empleando un emulador al que se le introducen como parámetros de entrada los valores de los indicadores de red (e. g., retardo, *jitter*, porcentaje de pérdidas de paquetes) extraídos de simulaciones de nivel de red y/o medidas de campo de una red LTE.

La maqueta propuesta consta de tres elementos: el servidor de *streaming*, el emulador de red IP móvil y el cliente de *streaming*. En los siguientes apartados se describen los formatos de vídeo 3D, los protocolos de red y los elementos que forman la maqueta.

A. Formatos de vídeo 3D

El formato de representación de vídeo 3D más simple es el formato CSV (*Conventional Stereo Video*). El formato CSV se compone de dos flujos de vídeo independientes (pero sincronizados), correspondientes al ojo izquierdo y derecho de un hipotético observador. Ambos flujos se pueden distribuir en paralelo utilizando sistemas convencionales, estrategia conocida como *Simulcast*. Este formato no requiere ningún procesamiento de geometría espacial, aportando sensación de profundidad al vídeo 2D, pero a costa de doblar la tasa de transferencia de un flujo de vídeo convencional. Por ello, se hace imperativa la utilización de métodos de compresión de la información para adaptarla a las limitadas capacidades de la red y los terminales móviles.

En la plataforma desarrollada, se utilizan los formatos 3D de tasa de transferencia reducida más habituales, que son SbS (*Side-by-Side*) y MVC (*Multi-View Coding*). El formato SbS es un formato compatible con los sistemas de distribución de vídeo tradicionales. Para ello, se aplica un proceso de diezmado en el dominio espacial de ambos ojos y las imágenes diezmadas se unen en un único fotograma convencional. El formato MVC, utilizado en el soporte de almacenamiento Blu-ray 3D, aplica codificación diferencial entre ambas vistas para eliminar la información redundante entre ellas [3].

B. Distribución del servicio de streaming de vídeo

Para la distribución de vídeo en redes IP, se pueden usar diferentes combinaciones de protocolos de aplicación, como HTTP (*Hyper-Text Transfer Protocol*) o RTP (*Real Time Protocol*), y protocolos de transporte, como TCP (*Transport Control Protocol*) o UDP (*User Datagram Protocol*). La combinación HTTP-TCP, con ciertas modificaciones sobre los estándares, es la escogida para el *streaming* de descarga progresiva, utilizado, p.ej., por YouTube. Por el contrario, RTP-UDP se usa para videoconferencia (p.ej., *Skype*), por ser más eficiente en el uso del ancho de banda.

La Fig. 1 muestra la arquitectura de protocolos del sistema de evaluación propuesto. En la figura, se identifican el cliente, el emulador de red y el servidor. Tanto el cliente como el servidor soportan las diferentes combinaciones de protocolos, aunque en este trabajo se considera la opción del servicio de *videostreaming* 3D sobre RTP/UDP. Por su parte,

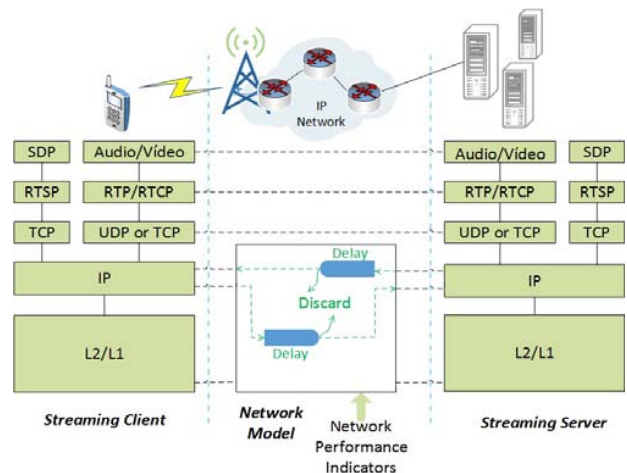


Fig. 1. Arquitectura de protocolos del sistema de evaluación.

el emulador de red introduce pérdidas y retardo sobre los paquetes, según su configuración de parámetros.

El servicio de *streaming* consiste en enviar los datos de forma continua estableciendo una conexión permanente entre cliente y servidor mediante RTSP (*Real-Time Streaming Protocol*). Una vez iniciada la sesión, normalmente los datos del vídeo y el audio serán enviados en sendas parejas de puertos RTP/RTCP (datos y control) sobre UDP, aunque también existe el mecanismo llamado *interleaving* que hace que RTSP se encargue del transporte de los datos.

Los indicadores de rendimiento principales para evaluar la calidad de la comunicación en servicios de vídeo son: velocidad de transferencia, retardo y *jitter*. La diferencia de retardo entre dos paquetes i y j , $D(i,j)$, y el *jitter* entre llegadas, $J(i)$, sufridos por cada uno de los flujos RTP se calculan a partir de las ecuaciones definidas en la RFC 3550,

$$D(i,j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \quad (1)$$

Donde S_i y S_j son los RTP *timestamp* de los paquetes i y j , y R_i y R_j son los instantes de llegada de los paquetes i y j en unidades de RTP *timestamp*, y

$$J(i) = J(i-1) + (|D(i-1,i)| - J(i-1))/16 \quad (2)$$

donde $J(i-1)$ es el *jitter* del paquete $i-1$ y $D(i-1,i)$ es la diferencia de tiempos de transmisión y recepción de los paquetes i e $i-1$ en orden de llegada (no necesariamente consecutivos).

C. Elementos de la maqueta

La Fig.2 muestra los elementos hardware que conforman la maqueta del sistema, compuesta por un servidor, un emulador de red y dos tipos de cliente. Como elementos software se han utilizado VLC, Wireshark y NetEm.

El servidor es un PC con procesador i5-750 a 2.67GHz y 4 GB de RAM con sistema operativo Ubuntu 14.04 LTS 64 bits y *kernel* 3.11.10.4. en el que se ha instalado el software VLC *media player* como servidor de *streaming*.

Un primer tipo de cliente es un PC con las mismas características que el anterior en el que también se ha instalado el sistema operativo Ubuntu 14.04 LTS 64 bits. En



Fig. 2. Componentes de la maqueta de evaluación.

él, se utiliza también el mismo reproductor multimedia VLC, pero ahora como cliente de *streaming*. Tanto en el servidor como en este cliente se ejecuta el software *Wireshark* como analizador de tráfico.

El otro tipo cliente es un terminal móvil modelo LG Optimus 3D-P920 con sistema operativo *Android 2.2.2*, en el que se ha instalado *VLC media player* para *Android v0.1.3*. Éste se emplea para las medidas de calidad subjetivas sobre pantalla pequeña. Estas medidas pueden realizarse en tiempo real haciendo uso de la conectividad Wifi del terminal, o en tiempo diferido reproduciendo los archivos almacenados en el cliente PC con las secuencias de vídeo deterioradas tras su paso por el emulador de red.

El emulador de red se instala en un PC con un procesador i5-3470 a 3.2GHz, 8 GB de RAM con sistema operativo Ubuntu 12.04 LTS 64 bits. Este ordenador consta de dos tarjetas de red para conectar al servidor y al cliente de *streaming* a través de él. La conexión entre el servidor de *streaming* y el emulador y entre este y el cliente se hace por medio de un cable de par trenzado cruzado para tener una red completamente aislada. Se han de configurar convenientemente las tablas de enrutamiento a través del comando de Linux *ip route*. El software que realiza las funciones de emulador es *NetEm*.

La versión de *VLC media player* elegida ha sido v2.1.4 *Rincewind*. Este reproductor está muy extendido por su gratuidad, facilidad de uso, variedad de codecs soportados, rapidez, fiabilidad y versatilidad. Es una plataforma de código libre, que soporta la mayoría de formatos multimedia, así como varios protocolos de *streaming*. En nuestro caso, además de para la reproducción de archivos multimedia, se ha hecho uso de la captura y almacenamiento de la secuencia de reproducción en el terminal cliente, para su posterior procesamiento y análisis comparativo del vídeo emitido con el recibido.

También *VLC* permite incluir fácilmente nuevos codecs de vídeo. Así que se ha ampliado esta herramienta y se han incluido las rutinas JMVC, que es el software de referencia del códec MVC desarrollado por el JVT (*Joint Video Team*) [2].

Como aplicación para la captura y análisis de tráfico se ha escogido *Wireshark v1.6.7-1* por ser el analizador de protocolos más utilizado por la comunidad científica para realizar análisis y diagnosticar problemas en redes de comunicaciones. Proporciona una funcionalidad similar a la de *tcpdump*, sin embargo añade una interfaz gráfica y muchas opciones de organización y filtrado de capturas que hacen que sea un entorno más amigable.

Como emulador de red se ha escogido *NetEm*, el paquete *iproute2 v3.10.0-1ubuntu1*. Este software es un emulador de red de código abierto ampliamente utilizado en sistemas de evaluación y desarrollo de protocolos de red y aplicaciones. Reproduce la dinámica de una red mediante la introducción de retardos, jitter y pérdidas, duplicando o eliminando paquetes.

A partir de estos elementos hardware y software se construye la maqueta que nos permitirá emular un sistema de *videostreaming* sobre una red móvil. Es necesario obtener los parámetros que nos permitan modelar una red LTE para introducirlos de forma adecuada en el emulador de red.

D. El emulador de Red: NetEm

NetEm[8] es un emulador de red que permite controlar el retardo, las pérdidas y la duplicación y el reordenamiento de los paquetes IP. A partir de la versión 2.6.8 (2.4.28), esta herramienta se encuentra integrada en el kernel de Linux y se controla a partir del comando 'tc' que es parte del paquete *iproute2*. *NetEm* se configura a través de este comando y puede automatizarse mediante el uso de scripts en shell para mayor comodidad.

En *NetEm*, la pérdida de paquetes se implementa descartando al azar un porcentaje de los paquetes antes de que se pongan en cola. Dicha pérdida se especifica en la interfaz de comando como un porcentaje. La emulación de pérdidas de paquetes a ráfagas se realiza mediante uno de los modelos de pérdidas disponibles (correlación, modelo de cuatro estados, modelo Gilbert-Eliot).

Los parámetros de retardo y su variación, *jitter*, se describen por el valor medio (μ), la desviación estándar (σ) y, opcionalmente, un valor de correlación (ρ). Por defecto, *NetEm* utiliza una distribución uniforme de retardo y jitter, que puede cambiarse a otra función de distribución (pareto, paretonormal o normal) u otras específicas a partir de datos experimentales. Esta última opción es la elegida para configurar convenientemente el emulador a partir de los datos obtenidos del simulador LTE.

Aunque *NetEm* es una herramienta con mucho potencial, tiene defectos en su comportamiento, como evidencian trabajos anteriores [9]. Estos estudios muestran que si a *NetEm* se le obliga a guardar la integridad del flujo de paquetes, para evitar su reordenamiento, se modifica el valor medio y la forma de la función de distribución del retardo. Este comportamiento hace que *NetEm*, en su configuración habitual, no sea apropiado para modelar sistemas de colas.

En este trabajo, para evitar este comportamiento, se modifica el código fuente de *NetEm*. En concreto, se modifica el método que se encarga de desencolar los paquetes (función *netem_enqueue*). El objetivo del cambio es modelar la acumulación de retardo que se produce cuando los paquetes llegan juntos al principio de cada cuadro.

Para ello, el retardo de transmisión de paquete se calcula como la suma de dos componentes, llamados retardo en cola y retardo de transmisión como primero en cola. El retardo en cola, D_q , es la diferencia de tiempo entre el instante de llegada del paquete y el instante en el que paquete pasa a ser el primero en la cola. El retardo de transmisión como primero en cola, D_{1st} , es la diferencia de tiempo entre el instante en el que el paquete pasa a ser el primero en la cola y el instante en que finalmente es servido.

```

static int netem_enqueue(structsk_buff *skb,
structQdisc *sch)
{
delay = tabledist(q->latency, q->jitter,
&q->delay_cor, q->delay_dist);

[...]

if (last) {
/*Last packet in queue is reference (now),
*calculate this time bonus and add
*from delay*/

delay += netem_skb_cb(last)->time_to_send - now;
delay = max_t(psched_tdiff_t, 0, delay);
now = netem_skb_cb(last)->time_to_send;
}
[...]
}

```

Fig. 3. Líneas de código añadidas en *NetEm*.

El retardo en cola de un paquete se calcula acumulando el retardo de transmisión de todos los paquetes que aún quedan por transmitirse en la cola. Por su parte, el retardo de transmisión primero en cola se genera de forma aleatoria a partir de la distribución de probabilidad obtenida en el simulador de red. Esta modificación en el código de *NetEm* evita el desordenamiento de los paquetes, haciendo que se transmitan según una cola con disciplina de servicio FIFO. Así se consigue la integridad del *videostreaming*, manteniendo los valores estadísticos y la forma de la función de distribución del retardo.

La Fig. 3 muestra las líneas de código que se introducen. En primer lugar, se obtiene el valor de retardo primero en cola a partir de la tabla que modela la distribución acumulada de probabilidad empírica (*Empirical Cumulative Density Function*, ECDF) del retardo (función *tabledist*). Después, si el paquete no encuentra la cola vacía (*last=1*), al retardo primero en cola se le añade el tiempo de espera en la cola (*netem_skb_cb(last)->time_to_send - now*). De esta manera, se consigue que no haya paquetes que se adelanten a otros debido a que su *delay* sea menor.

Para introducir la función de probabilidad de retardo obtenida desde el simulador LTE en el emulador, en la función *tabledist* ha sido necesario modificar el tipo de datos que soportaba inicialmente *NetEm*.

III. SIMULACIÓN DE INTERFAZ RADIO LTE

Para configurar los parámetros del emulador de red con valores realistas, se realiza una campaña de simulaciones con un simulador LTE de nivel de sistema [10]. El objetivo de estas simulaciones es estimar el rendimiento de la transmisión de paquetes en la red de acceso con distintas condiciones radio.

La Fig.4 muestra el escenario, con 53 celdas con antena omnidireccionales situadas en el centro de cada celda, distribuidas de forma regular en el espacio. En dicho escenario, se distribuyen varios usuarios por celda del servicio de *videostreaming* 3D. La tabla 1 describe los parámetros del simulador. Por simplicidad, se simula solo el enlace descendente, que es el principal cuello de botella en la descarga de secuencias de vídeo de un servidor.

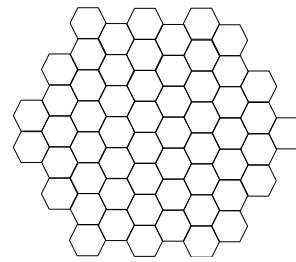


Fig. 4. Escenario de simulación.

Tabla 1. Parámetros del simulador de red LTE.

Frecuencia de portadora	2.5 GHz
Ancho de banda del sistema	10 MHz (50 Physical Resource Blocks)
Número de celdas	57
Radio de celda	1.5 km
Modelo de propagación	Hata
	Desvanecimiento lento lognormal 6dB
Potencia isotrópica radiada efectiva	43 dBm
Densidad de potencia de ruido	-174 dBm/Hz
SchedulereNB	Proportional Fair
Esquema de transmisión MIMO	2x2Beamforming
Modelo de canal	Extended pedestrian A
Método decodificación de canal	Soft-Output Viterbi Algorithm
BLER objetivo	10%
Modelo de servicio	Traza de vídeo real H.264
Tiempo de red simulado	40s
Resolución temporal	1ms (TTI)

La salida principal obtenida del simulador es una traza con los retardos de transmisión de los paquetes de nivel IP en el enlace descendente obtenidas por un usuario en unas condiciones de red determinadas. Se distinguen 4 condiciones diferentes de trabajo, que pretenden reflejar la experiencia de un usuario cuando se sitúa en el centro o en el borde de la celda, en condiciones de baja o alta carga de red. La posición del usuario en la celda determina la calidad de señal (relación señal a interferencia más ruido) experimentada por el usuario, mientras que las condiciones de carga de red determinan el número de recursos radio disponible para el usuario. En los experimentos, la carga de red se controla modificando el número de usuarios por celda entre 2 usuarios (para el que se obtiene un porcentaje de ocupación de bloques de recursos físicos de un 25%) y 5 usuarios (para el que se obtiene un porcentaje del 75%). La posición de centro o borde celda se consigue fijando la posición en el escenario del usuario del que se obtiene la traza de retardo.

La Fig.5 presenta la ECDF del retardo de transmisión de paquete para los 4 casos considerados, mientras que la tabla 2 detalla algunos estadísticos de la transmisión de paquetes. Se aprecia cómo la mediana del retardo medio de transmisión de paquete oscila desde 8 ms para el caso más favorable (usuario centro de celda y carga baja) hasta valores por encima de los 200 ms para el caso más desfavorable (usuario borde celda y carga alta).

Para cada caso de uso, la traza de retardo de transmisión de paquete obtenida se procesa para obtener la traza correspondiente de retardo de paquete cuando el paquete es el primero en la cola. Concretamente, el retardo de paquete primero en cola, D_{1st} , se estima a partir del retardo total de transmisión, D , como

$$D_{1st}(i) = D(i) - D_q(i) = D(i) - \max(T(i-1) - A(i), 0) \quad (3)$$

donde D_q es el retardo de espera en cola, estimado a partir de la diferencia entre el instante de transmisión del paquete anterior, $T(i-1)$, y el instante de llegada del paquete en cuestión, $A(i)$.

Los valores de las componentes de retardo mostradas en la tabla 2 ponen de manifiesto que la mayor parte del retardo se debe al retardo de encolado. Un análisis detallado muestra que este comportamiento se debe a la llegada simultánea de paquetes al principio de cada cuadro. La Fig. 6 muestra un ejemplo de la evolución del tiempo entre llegadas de paquetes para una traza de vídeo. En ella, se aprecia que, tras un periodo inicial de intercambio de paquetes de señalización para establecer la sesión RTP, los paquetes que encapsulan el vídeo se generan en ráfagas que se repiten con periodo de cuadro (40 ms). En la figura se aprecia también que el número de paquetes por ráfaga mantiene un cierto patrón repetitivo, del que se puede deducir la estructura de grupo de imágenes del compresor, formada por 1 cuadro tipo I (*Intracuadro*) por cada 7 cuadros de tipo P (*Predictivo*). De la distancia entre picos de la serie de diferencias de instantes de llegada, se puede deducir que el número medio de paquetes que forman los cuadros I y P son 35 y 8, respectivamente.

A partir de las trazas de retardo obtenidas del simulador LTE se obtiene el retardo en cola que puede observarse en la Fig. 7. Este retardo presenta una curva en diente de sierra semiperiódico que muestra como al inicio de cada cuadro se produce un incremento drástico de este tiempo que va decreyéndose a medida que se sirven los paquetes.

También a partir de las trazas de retardo obtenidas del simulador LTE, se construye la ECDF del retardo de paquete primero en cola, D_{1st} . En la Fig. 8 se muestra la ECDF del retardo de paquete primero en cola para cada uno de los casos de usos. Se aprecia cómo la mediana de D_{1st} oscila entre 1,17 ms para el caso más favorable (usuario centro de celda y carga baja) hasta 6,5 ms para el caso más desfavorable (usuario borde celda y carga alta).

Con cada una de las ECDFs del retardo de paquete primero en cola de la Fig. 8, se genera una realización de 4096 valores aleatorios de retardo. Dichos valores, una vez convertidos a unidades de *tick* (1 *tick* = 64 μ s) y ordenados de menor a mayor, se introducen en la tabla de 512*8 valores utilizada por *Netem* para generar los retardos de paquete primero en cola en cada caso de uso.

Tabla 2. Estadísticos de transmisión de paquetes.

Carga de red	Baja		Alta	
	Centro	Borde	Centro	Borde
Retardo total medio [ms]	14.96	48.80	30.90	4542.9
Retardo en cola medio [ms]	12.48	45.77	28.70	4258.8
Retardo primero en cola medio [ms]	1.20	3.03	2.48	284.1
Jitter medio [ms]	3.50	5.18	3.81	2.79

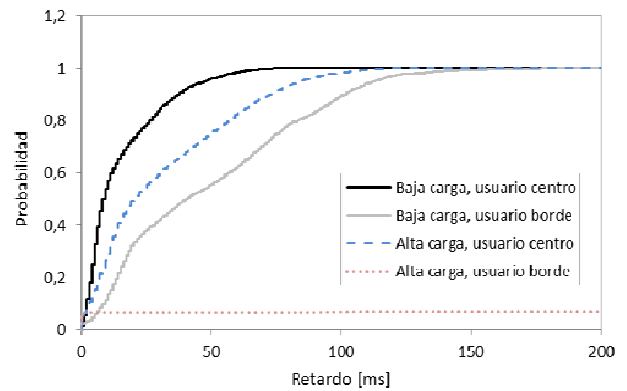


Fig. 5. Distribución acumulada de probabilidad del retardo de paquete total en distintas condiciones.

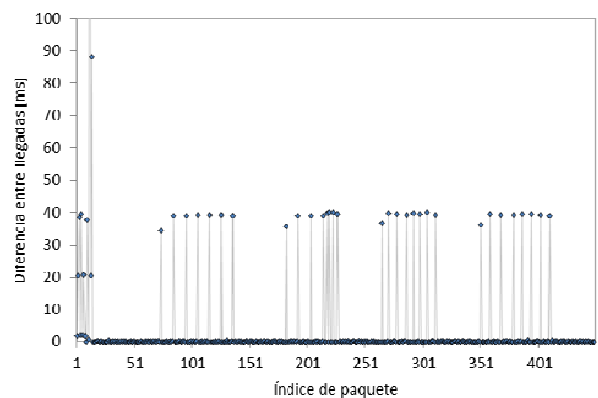


Fig. 6. Ejemplo de evolución del tiempo entre llegadas de paquetes.

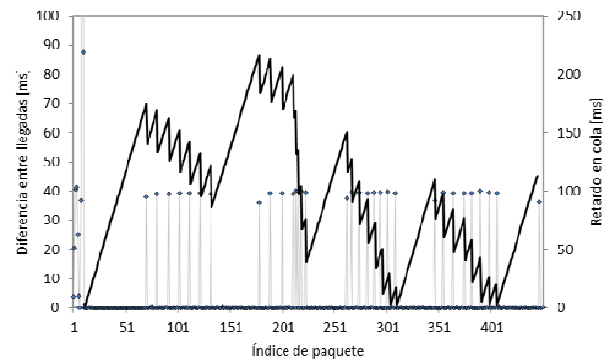


Fig. 7. Ejemplo de evolución del tiempo entre llegadas de paquetes y el tiempo de espera en cola que experimenta el usuario.

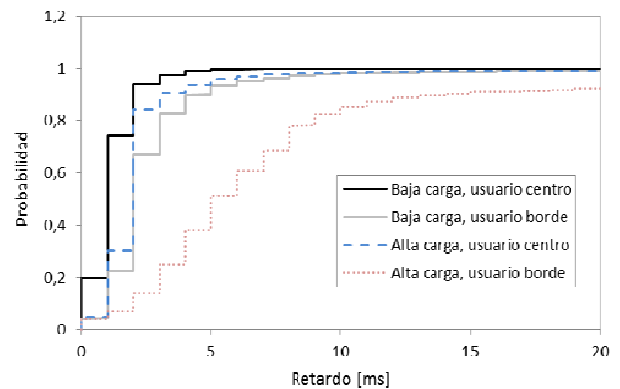


Fig. 8. Distribución acumulada de probabilidad del retardo de paquete primero en cola en distintas condiciones.

IV. METODOLOGÍA EXPERIMENTAL

Para validar el funcionamiento de la plataforma, primero se configuran los elementos de la maqueta según los resultados del simulador LTE, luego se lleva cabo la distribución de secuencias 3D, y por último se procesan las medidas de retardo obtenidas.

La configuración de la maqueta consiste fundamentalmente en la selección de la secuencia de vídeo a distribuir, la selección de la tabla que modela la ECDF de retardo primero en cola en el emulador de red, y la configuración del tamaño del buffer del reproductor.

Antes de comenzar con la conexión entre cliente y servidor, se debe comprobar la sincronización de los equipos mediante el protocolo NTP (*Network Time Protocol*). Esta fase es muy importante, ya que de ello depende la bondad de los resultados. En la Fig. 9 puede observarse cómo la falta de sincronismo entre los equipos produce una deriva considerable, en este caso del orden de 1.43 ms por cada 1000 paquetes. En la Fig. 10 se observa el resultado de la misma simulación una vez que servidor y cliente han sido previamente sincronizados.

Como la frecuencia de trabajo de *NetEm* es directamente la marcada por el *kernel*, se ha modificado el tiempo de reloj del *kernel* a 1 ms para mejorar la resolución temporal de *NetEm*. Con el mismo fin, y para mejorar el sincronismo entre los tres PCs, se han habilitado los temporizadores de alta resolución del sistema operativo.

Una vez sincronizados los equipos, se lanzan los analizadores de tráfico en ambos equipos, emisor y receptor. Una vez activadas las herramientas de monitorización, se inicia la transmisión desde el servidor *VLC streaming Server* al cliente *VLC media player* de la secuencia de vídeo seleccionada en el cliente. Al concluir la reproducción, los datos capturados por los analizadores de tráfico del emisor y receptor se almacenan para su posterior análisis.

En la tabla 3 se detallan las características de las secuencias de vídeo empleadas y los valores de los parámetros de configuración de *NetEm* y *VLC*. Todas las secuencias de vídeo empleadas en las pruebas están codificadas en formato SbS. El tamaño del buffer se ha elegido de 50 ms, que es apropiado para un servicio interactivo de voz y vídeo cuyo retardo extremo a extremo esperado no debería ser mayor de unos 200ms.

Para emular la red LTE, se configura el emulador de red *NetEm* de acuerdo a los cuatro escenarios que se han descrito en la sección anterior (alta/baja carga, usuario centro/borde). Para ello, se indica a *NetEm* la tabla concreta que modela el retardo primero en cola, generada a partir de una de las distribuciones de la Fig. 8.

Las trazas de retardo del cliente y servidor se capturan mediante *Wireshark*. Con su funcionalidad *RTP Analysis*, en teoría es posible medir el retardo medio, el *jitter* y las pérdidas de paquetes de un flujo de datos RTP. Sin embargo, durante los experimentos se ha comprobado que *Wireshark* no muestra los resultados de *jitter* y retardo medio de forma adecuada como cabría esperar, ya que no es capaz de interpretar correctamente la decodificación de flujos RTP. El retardo detectado es leído dentro de la *ExpertInfo* de *Wireshark*. Las estadísticas del *jitter* sufrido por el tráfico RTP se han calculado mediante la fórmula (2), a partir de los tiempos de transmisión, *timestamp* de la cabecera RTP del

paquete, y los tiempos de recepción de los paquetes capturados por el analizador del equipo receptor.

La calidad de experiencia ofrecida al observador final puede comprobarse reproduciendo las secuencias de vídeo recibidas en la pantalla del reproductor *VLC* del cliente. Para tener una medida objetiva del impacto de las fluctuaciones del retardo sobre la calidad del vídeo, se utiliza la funcionalidad de *VLC* que permite grabar las secuencias recibidas. Estas secuencias, una vez almacenadas, se procesan con el reproductor *ffmpeg* [11], que permite evaluar la calidad de la secuencia en función del número de cuadros perdidos y duplicados.

Tabla 3. Parámetros de configuración de las pruebas.

Nombre de secuencia	pd01	pd02	pd03	pd04
Codec de vídeo	H.264/AVC 3086 kbps	H.264/AVC 2207 kbps	H.264/AVC 4385 kbps	H.264/AVC 1729 kbps
Duración [s]	48	23	40	60
Parámetros NetEm	[ECDF para cada uno de los 4 casos de uso]			
Parámetros VLC	Tamaño de buffer: 50ms			

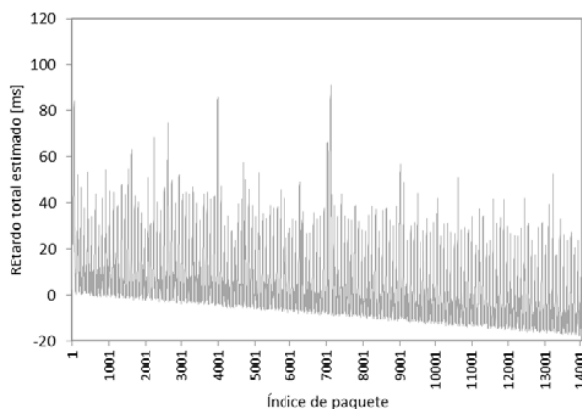


Fig. 9. Retardo extremo a extremo para cada paquete cuando el cliente y servidor no están sincronizados.

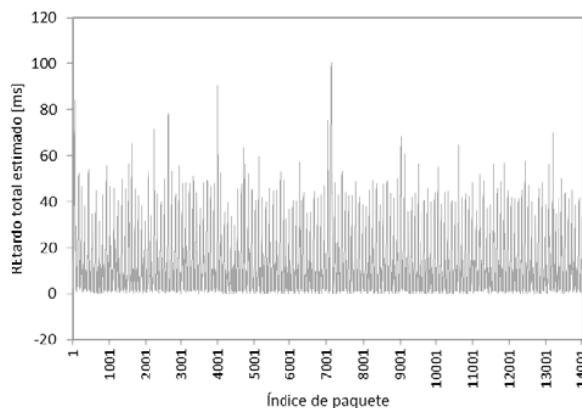


Fig. 10. Retardo extremo a extremo para cada paquete cuando el cliente y servidor están sincronizados.

V. RESULTADOS

En esta sección se muestran los resultados a los experimentos realizados para los cuatro casos de estudio que modelan las distintas condiciones que puede experimentar un

usuario, situado en el centro o en el borde de la celda, en condiciones de baja o alta carga de red. Por brevedad, solo se muestran los resultados obtenidos para una secuencia de vídeo (secuencia *pd01*, cuyas características se describen en la tabla 3). En la Fig. 11 se muestra un cuadro de la secuencia de vídeo escogida, donde se aprecian los fotogramas de la vista izquierda y derecha. En ella, se observa cómo esta secuencia consta de una única toma con poco movimiento, en la que se ubica una persona en primer plano gesticulando delante de un fondo estático.

En la Fig. 12 se muestra la función de distribución acumulada del retardo extremo a extremo obtenida mediante el emulador de red (*Netem*), comparada con la del simulador LTE, para cada uno de los cuatro casos. En la figura, la configuración del eje *x* ha sido seleccionada para apreciar con detalle las diferencias entre las curvas de menor retardo. Nótese que en el caso (carga alta, usuario de borde), al ser un sistema de cola inestable (es decir, tasa de llegada paquetes mayor que tasa de salida de paquetes), el retardo es muy elevado. Por ello, la curva correspondiente está desplazada en el eje *x*, no siendo visible. De la figura, puede concluirse que las curvas del simulador y el emulador se ajustan casi perfectamente para los casos (baja carga, usuario de centro) y (alta carga, usuario de centro), mientras que el ajuste es menos preciso, aunque aceptable, en el caso de (baja carga, usuario de borde).

La tabla 4 resume los principales estadísticos de la transmisión de paquetes. Puede observarse otra vez el buen ajuste entre los valores obtenidos mediante el emulador y el simulador para la mayoría de los escenarios. En el caso (baja carga, usuario de centro), el error relativo en el retardo medio es inferior al 0.03%. En el caso (alta carga, usuario de centro), es del 4%, mientras que (baja carga, usuario de borde), el error relativo en el retardo medio es del 10%. Estos resultados son acordes a lo observado en la Fig. 11. En el caso (alta carga, usuario de borde), donde el sistema es inestable, la desviación en el retardo medio es considerable.

En la tabla 4, también se observa que el *jitter*, en todos los casos, tiene un valor del orden del retardo de paquete primero en cola. Como cabría esperar, la diferencia del retardo entre dos paquetes consecutivos de la misma ráfaga (cuadro de vídeo) se aproxima al retardo medio de transmisión primero en cola, tal como se deduce de la Fig. 7. Comparando los valores del simulador y el emulador, se deduce que el error relativo en el *jitter* está por debajo del 5% en todos los casos donde el sistema de cola es estable.

En la Tabla 5 se presentan diversos estadísticos de rendimiento asociados a la decodificación del vídeo. Se observa cómo, en los tres casos donde el sistema es estable, el número de cuadros duplicados o perdidos, es muy inferior al obtenido cuando el sistema es inestable. Aun así, de las cifras de la tabla no se pueden extraer conclusiones relevantes sobre cuál de los tres casos es más favorable, dado que las cifras de cuadros perdidos/duplicados son muy pequeñas en cualquiera de los casos, lo que las hace dependientes de la realización específica de las variables aleatorias del experimento.



Fig. 11. Imagen perteneciente a la secuencia *Bullinger* (secuencia *pd0*).

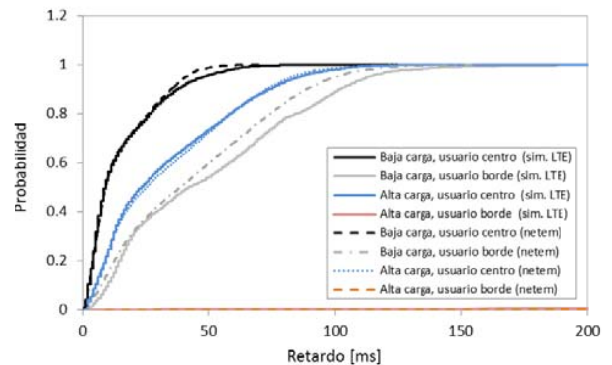


Fig. 12. Distribución acumulada de probabilidad del retardo total de paquete en el emulador y el simulador.

Tabla 4. Estadísticos de transmisión de paquetes.

Carga de red	Baja		Alta	
	Centro	Borde	Centro	Borde
Retardo total medio (sim. LTE) [ms]	15.16	48.91	31.77	4857
Retardo total medio (NetEm) [ms]	14.19	43.71	33.06	21278
Jitter medio (sim. LTE) [ms]	2.09	4.42	3.86	9.86
Jitter medio (NetEm) [ms]	2.00	4.53	3.77	6.43

Tabla 5. Estadísticos de reproducción del vídeo.

Carga de red	Baja		Alta	
	Centro	Borde	Centro	Borde
Nº cuadros duplicados descompresor	16	12	46	585
Nº cuadros perdidos descompresor	3	0	0	6

VI. CONCLUSIONES

En este trabajo se ha presentado una maqueta de sistema para la evaluación de la calidad de experiencia del servicio de *videostreaming* 3D en una red LTE. Las aplicaciones que conforman la maqueta son *VLC* como servidor y cliente de *videostreaming*, *NetEm* como emulador de red y *Wireshark* como herramienta de monitorización.

Para modelar el retardo de transmisión de forma precisa, se ha modificado el código fuente de *NetEm* para implementar un sistema de colas, donde el tiempo de respuesta del sistema es la suma del retardo de encolado y el retardo una vez que el paquete es primero en la cola. Esta última componente del retardo se obtiene a partir de las distribuciones de retardo de paquete obtenidas en un simulador LTE de nivel de sistema configurado para modelar escenarios realistas.

Los resultados han demostrado que, a diferencia del emulador original, la maqueta permite modelar de forma bastante precisa el retardo medio y el jitter bajo diferentes condiciones radio. Concretamente, en los casos donde el sistema es estable (es decir, cuando la tasa de transmisión de datos media supera a la tasa media de codificación del vídeo), el error relativo del retardo y el jitter respecto al simulador LTE es inferior al 10% y el 5%, respectivamente.

Se tiene previsto extender el presente estudio con la realización de una campaña de medidas de la calidad subjetiva experimentada cuando las secuencias se reproducen con un terminal móvil con pantalla 3D. Asimismo, se ampliará la maqueta con nuevos formatos de codificación de vídeo (p.ej., vídeo y profundidad) y métodos de medida de calidad objetiva de vídeo 3D (p.ej., PSNR, MS-SSIM).

AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad (proyecto TIN2012-36455) y la Junta de Andalucía (P10-TIC-6897). Asimismo, los autores quieren agradecer la colaboración de Almudena Sánchez Martínez y Francisco Blánquez Casado en la realización de los experimentos.

REFERENCIAS

- [1] I. Kulik, T. A. Trinh, "Investigation of Quality of Experience for 3D Streams in Gigabit Passive Optical Network", EUNICE 2011, LNCS 6955, pp. 157-168, 2011.
- [2] Z. Liu, et al., "Experimental Evaluation of H.264/Multiview Video Coding over IP Networks", ISSC 2011, Jun 2011.
- [3] C. Jean, et al., "QoE-based Packet Drop Control for 3D-Video Streaming over Wireless Networks", LANC'12, pp. 59-66, Oct 2012.
- [4] S. Singh, et al. "Video Capacity and QoE Enhancements over LTE", ICC 2012, pp. 7071 - 7076, 2012.
- [5] I. Palomo, M. Solera, M. Toril, J. Poncela-González, G. Gómez "Maqueta de sistema para la evaluación del servicio de videostreaming 3D sobre LTE", URSI 2014, Valencia (España), Sep. 2014.
- [6] G. Gómez, et al., "A Real-Time End-to-End Testbed for Evaluating the Performance of Multimedia Services", FMN 2009, LNCS 5630, pp. 212-217, 2009.
- [7] P. Uppu, S. Kadimpati, "QoE of Video Streaming over LTE Network", MS Thesis, Oct 2013.
- [8] <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>.
- [9] J. Sliwinski, A. Beben, P. Krawiec, "EmPath: Tool to Emulate Packet Transfer Characteristics in IPNetwork", Lecture Notes In Computer Science, Springer, nr 6003, 2010.
- [10] G. Gomez, D. Morales-Jimenez, J. J. Sanchez-Sanchez, J. Tomas Entrambasaguas, "A Next Generation Wireless Simulator based on MIMO-OFDM: LTE Case Study", EURASIP Journal on Wireless Communications and Networking, Vol. 2010, 2010.
- [11] <https://www.ffmpeg.org/>.

Problemas de seguridad y amenazas que pueden afectar al cloud híbrido de FINESCE

Julia Sánchez, Guiomar Corral, Jaume Abella, Agustín Zaballos

Departamento de Ingeniería
La Salle – Universidad Ramon Llull (URL)
Barcelona

{julias, guiomar, jaumea, zaballos}@salleurl.edu

Resumen- El proyecto FINESCE, enmarcado dentro del séptimo programa marco (FP7) de la Comisión Europea, pretende dar solución al caso de uso *Smart Energy* definiendo una infraestructura abierta basada en las TIC utilizada para desarrollar nuevas soluciones y aplicaciones en todos los campos del Internet del Futuro relacionados con el sector energético. Para ello, se propone utilizar un entorno basado en *Cloud Computing* que aportará una gran escalabilidad, rápido aprovisionamiento, capacidad de recuperación y eficiencia en costes, facilitando el desarrollo de aplicaciones y servicios para las compañías eléctricas. Este documento recoge las principales características de los modelos de entrega de servicios IaaS, SaaS y PaaS, y también del modelo de despliegue de *cloud híbrido*, para analizar sus problemas de seguridad y amenazas y poder justificar la elección del entorno de *cloud* que más se ajusta al caso de uso *Smart Energy* destacando sus flaquezas.

Palabras Clave- Cloud Computing, IaaS, SaaS, PaaS, amenazas en el cloud, seguridad en el cloud, jitel, telemática

I. INTRODUCCIÓN

Durante la última década, las *Smart Grids* han impulsado la revolución de la red eléctrica convirtiéndola en un conjunto de procesos automatizados y eficientemente controlados gracias a la incorporación de las TIC (*Tecnologías de la Información y las Comunicaciones*).

El proyecto FINESCE¹ tiene como objetivo principal contribuir al desarrollo de una infraestructura abierta basada en las TIC utilizada para desarrollar nuevas soluciones y aplicaciones en todos los campos del Internet del Futuro relacionados con el sector energético. Mediante la colaboración de La Salle R&D, el proyecto FINESCE incorporará el concepto de subestación virtual IEC61850 a través de prototipos IDEV desarrollados previamente en el proyecto INTEGRIS², los cuales se interconectarán a través del *cloud* de FIWARE³ utilizando diversos *Generic Enablers*, dando lugar a un sistema de almacenamiento distribuido basado en un *cloud híbrido*. De esta manera, se plantea una solución alternativa para las eléctricas en la cual dispondrán de un sistema de almacenamiento más robusto y

escalable gracias a la combinación del *Cloud Computing* con sus infraestructuras privadas de CPD (Centro de Proceso de datos).

Dado que este sistema se apoyará en tecnologías basadas en *Cloud Computing* aprovechando los costes reducidos, flexibilidad y escalabilidad de las soluciones que proporciona, resulta necesario estudiar sus implicaciones, sobre todo por lo que respecta a la seguridad y privacidad de los datos según la solución de *cloud* escogida, ya sean datos almacenados o en tránsito.

El NIST [1] define el *Cloud Computing* como un modelo que permite el acceso bajo demanda a un conjunto de recursos informáticos configurables compartidos que pueden aprovisionarse rápidamente y liberarse con un mínimo esfuerzo de gestión o interacción por parte del proveedor de servicios de *cloud* (CSP, *Cloud Service Provider*). Este aprovisionamiento se lleva a cabo mediante técnicas de virtualización con el fin de proporcionar una forma eficiente de entregar los recursos al momento a través de Internet. Esto permite implementar un modelo de negocio de pago por uso[2], en el cuál los clientes pueden elegir específicamente todos los recursos (p. ej., CPU, memoria, ancho de banda, políticas de seguridad y carga de hardware) que requieren, reduciendo los costes sin necesidad de invertir en soluciones TI internas. Por lo tanto, el cliente consigue un coste de mantenimiento muy reducido sin tener que preocuparse por cuestiones técnicas de gestión y mantenimiento de los recursos originales, pudiendo acceder a éstos desde cualquier lugar en cualquier momento, dado que el responsable de la disponibilidad de los recursos es el CSP [3].

Las soluciones de *cloud* ofrecen varios beneficios [4] como un rápido despliegue, pago por uso, reducción de costes, escalabilidad, rápido aprovisionamiento, flexibilidad, acceso a la red ubicua, mayor capacidad de recuperación (infraestructura del CSP muy protegida), protección del *hypervisor* contra ataques de red, recuperación de desastres y soluciones de almacenamiento de datos a bajo coste, controles de seguridad bajo demanda, detección en tiempo real de la manipulación del sistema y rápida reconstitución de los servicios. Sin embargo, el hecho de que la gestión de los datos físicos y máquinas se delegue a los CSPs, conservando el cliente un mínimo control sobre las máquinas virtuales, crea cierta inquietud y desconfianza [5]. ¿Cómo saben los clientes del *cloud* que su información no está teniendo ningún problema de disponibilidad y seguridad? ¿Está almacenada de manera segura su información?

¹ FINESCE. Future INternet Smart Utility ServiCEs. Proyecto encargado del caso de uso de *Smart Energy* en la segunda fase del programa FI-PPP respaldado por la Comisión Europea dentro del FP7 (<http://www.finesce.eu/>)

² INTEGRIS. INTElligent Electrical GRId Sensor communications. Infraestructura TIC flexible basada en un sistema de comunicación híbrido que integra tecnología PLC y Wireless (<http://fp7integriss.eu/>)

³ FIWARE. Plataforma basada en *Generic Enablers* (APIs abiertas de propósito general) para agilizar la tarea de los desarrolladores de aplicaciones del Internet del Futuro (<https://www.fiware.org/>)

No se puede perder de vista que las soluciones de *Cloud Computing* mueven el software de aplicación y las bases de datos de los clientes a grandes CPDs donde la gestión y los servicios no son de la misma confianza que cuando se alojan en una infraestructura interna. Este paradigma supone nuevos retos de seguridad que incluyen, aunque no se limitan a, vulnerabilidades de accesibilidad, vulnerabilidades de virtualización, vulnerabilidades de aplicaciones Web y de bases de datos, problemas de acceso físico, privacidad y problemas de control al depender de infraestructuras de terceros, problemas de identidad y administración de credenciales, cuestiones relacionadas con la verificación, la manipulación, la integridad, la confidencialidad, la pérdida y el robo de datos, cuestiones relacionadas con la autenticación del dispositivo o recurso demandado y suplantación de IP [4].

Este documento pretende recoger los requerimientos de seguridad básicos en el despliegue de una solución basada en *Cloud Computing* así como destacar ataques y vulnerabilidades a considerar para la aplicación de seguridad en el caso de uso *Smart Energy* del proyecto FINESCE. El documento está organizado de la siguiente manera. En la Sección II, se presentan los diferentes modelos de despliegue y entrega de servicios en entornos de *Cloud Computing* para extraer los requerimientos básicos de seguridad necesarios en el *cloud híbrido* de FINESCE. En la Sección III, se analizan las amenazas más relevantes que afectan al *cloud*. En la Sección IV, se analizan los problemas y vulnerabilidades según cada modelo de entrega de servicio. En la Sección V, se expone la solución propuesta para el proyecto FINESCE y sus implicaciones de seguridad destacando algunas vulnerabilidades del sistema. Finalmente, en la Sección VI se exponen las conclusiones del trabajo realizado.

II. REQUERIMIENTOS DE SEGURIDAD DE LOS MODELOS DE DESPLIEGUE Y ENTREGA DE SERVICIOS EN EL CLOUD

En un entorno de *Cloud Computing* existen diversos riesgos de seguridad según cómo los CSPs entreguen sus servicios a los clientes. Como se aprecia en la Fig. 1, se debe considerar de manera transversal la resolución de los riesgos asociados a (1) la seguridad del almacenamiento de datos, (2) la seguridad en la transmisión de los datos, (3) la seguridad de las aplicaciones y (4) la seguridad relacionada con recursos de terceros. Cada modelo de entrega de servicios *cloud* (IaaS, PaaS o SaaS) proporcionará un conjunto de recursos de manera transparente con las siguientes características [1]:

- **Autoservicio bajo demanda.** Cualquiera puede aprovisionar y consumir recursos por su cuenta.
- **Acceso a red ubicua.** Acceso a los recursos del *cloud* a través de redes de servicios comunes como Internet.
- **Elasticidad rápida.** Capacidad de escalar casi al instante si se aumenta la necesidad de recursos.
- **Servicio cuantificado.** Seguimiento del consumo de recursos a fin de poder contabilizar los costes del modelo “pago por uso”.
- **Multi-tenancy.** Una sola instancia de una aplicación software sirve para múltiples clientes o *tenants*.

Además, los modelos de entrega de servicios pueden proporcionarse a través de diversos modelos de despliegue

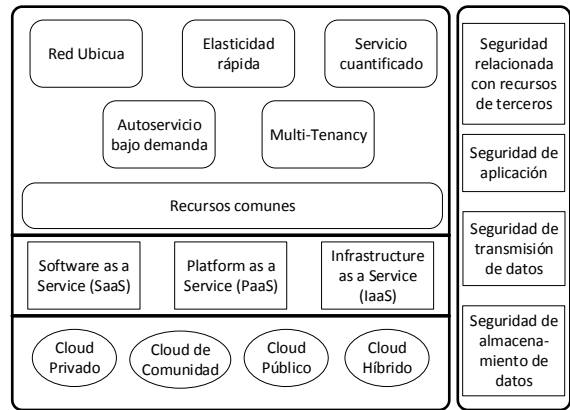


Fig. 1. Complejidad de la seguridad en un entorno *cloud* [4]

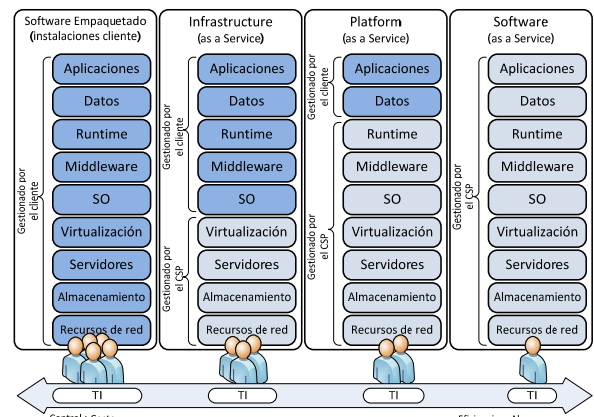


Fig. 2. Modelos de entrega de servicios en el *cloud*

(*cloud* privado, público, híbrido y de comunidad) que también tienen sus riesgos de seguridad por naturaleza [2].

A. Modelos de Entrega de Servicios

Como se aprecia en la Fig. 2, existe un compromiso entre el control del sistema, los datos y la eficiencia en costes según el modelo de entrega empleado. Cuanto menos control existe por parte del cliente, menores son los costes de implementación de las aplicaciones de negocio. Esto implica una pérdida de confianza debido a que la seguridad depende en gran parte del CSP. De todas maneras, el cliente está obligado a confiar en que la seguridad se extiende a lo largo de toda la pila ya que no le queda otra alternativa. Para definir las condiciones de entrega de servicio y aumentar esta confianza, se establecen SLA (*Service Level Agreements*) entre los clientes y sus proveedores asegurando la calidad, disponibilidad, fiabilidad y rendimiento de los recursos entregados [6].

Infrastructure-as-a-Service (IaaS) abstrae por completo el hardware subyacente asignando al cliente recursos físicos bajo demanda (típicamente en un entorno de virtualización), ofreciendo almacenamiento de datos, redes o capacidades computacionales y permitiendo que los usuarios utilicen la infraestructura como un servicio. IaaS sólo proporciona seguridad básica [2], incluyendo defensas perimetrales, tales como firewalls, sistemas de prevención de intrusiones (IPS) y sistemas de detección de intrusiones (IDS). También incluye balanceo de carga para proporcionar más disponibilidad y VMMs (*Virtual Machine Monitors*) para monitorizar el

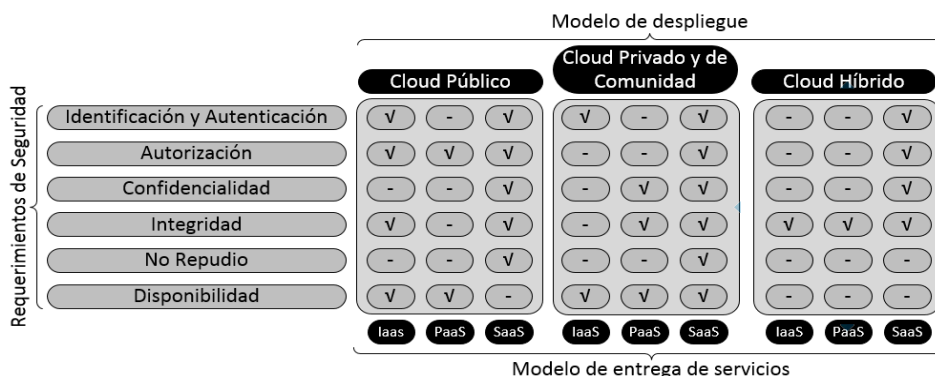


Fig. 3. Requerimientos de seguridad según modelo de entrega de servicios y para los diferentes modelos de despliegue [2]. El check (✓) significa que es un requerimiento obligatorio mientras que (-) significa que es un requerimiento opcional

comportamiento de las máquinas virtuales y proporcionar aislamiento entre ellas.

Platform-as-a-Service (PaaS) permite a los clientes construir sus propias aplicaciones mediante la entrega de un conjunto de herramientas y plataformas de desarrollo que ofrecen un ciclo de vida completo sin tener que preocuparse por el hardware y software que lo soporta. Algunos ejemplos de PaaS son Google App Engine, IBM SmartCloud y CloudHub. El *cloud* alberga arquitecturas SOA (*Software Oriented Architecture*) para web que ocultan los elementos subyacentes. Por este motivo, y dado que los atacantes son propensos a atacar código visible, es necesario disponer de un conjunto de métricas de codificación de seguridad para cuantificar la calidad del código escrito y evitar el desarrollo de aplicaciones expuestas a ataques [2].

Software-as-a-Service (SaaS) entrega aplicaciones software como servicio, como CRM y ERP, a través de Internet sin necesidad de instalar programas. Este modelo mejora la eficiencia operativa y también reduce los costes para los clientes. Se conocen muchos problemas de seguridad relacionados con los componentes básicos de las aplicaciones SaaS [7]. Desde la perspectiva del cliente, es difícil de entender si los datos están bien asegurados y si las aplicaciones estarán disponibles en todo momento debido a la falta de visibilidad sobre la forma en que los datos se almacenan y se implementan las aplicaciones. Los retos [2] en este modelo se centran en cómo preservar o mejorar la seguridad anteriormente proporcionada por los sistemas de *hosting* tradicionales.

B. Modelos de Despliegue

Como se aprecia en la Fig .2, los modelos de entrega de servicios pueden proporcionarse a través de diferentes modelos de despliegue en el *cloud*: (1) *Público*, (2) *Privado*, (3) *Híbrido*, (4) *De Comunidad*. En [2][8] se recogen las características de cada modelo de despliegue.

Para el caso de uso *Smart Energy*, interesa analizar las características relacionadas con el *cloud híbrido* las cuáles se detallan a continuación:

- Combinación de dos o más métodos de despliegue (comúnmente público y privado).
- Gestión a cargo de la organización o terceros.
- Disponibilidad de recursos que pueden estar dentro o fuera de las instalaciones del cliente.

- Acceso a través de Internet a múltiples entidades bien definidas pero limitadas.
- Más seguro que el *cloud* público ya que no todo el sistema depende del CSP pero menos seguro que el *cloud* privado dado que parte de la infraestructura de CPD está fuera del control del cliente y resulta difícil gestionar e identificar responsabilidades de seguridad cuando surge algún inconveniente.

La Fig. 3 resume los requerimientos de seguridad básicos de cada modelo de entrega de servicios según el modelo de despliegue sobre el que se asiente. Estos requerimientos son: (1) *Identificación y Autenticación*, (2) *Autorización*, (3) *Confidencialidad*, (4) *Integridad*, (5) *No repudio*, (6) *Disponibilidad*.

Cada modelo de despliegue presenta sus propios problemas y cuestiones de seguridad específicas. Los clientes de CSPs y organizaciones deben tener en cuenta varios factores, entre ellos, el presupuesto disponible, el propósito de los requerimientos del *cloud* y la seguridad, antes de decidir sobre un modelo específico.

Los requerimientos básicos de seguridad a tener en cuenta en el *cloud híbrido* desplegado para el caso de uso *Smart Energy* de FINESCE se reducen a: *Integridad* para IaaS y PaaS y *Autorización, Confidencialidad, Integridad e Identificación y Autenticación* para SaaS.

III. AMENAZAS DE SEGURIDAD EN EL CLOUD

El mayor problema al cual se enfrenta el *Cloud Computing* es asegurar la **confidencialidad** e **integridad** de los datos y la **disponibilidad** de servicios. Un componente central para gestionar los riesgos asociados a este problema es entender la naturaleza de las amenazas de seguridad en el *cloud*. En [9] la CSA (*Cloud Security Alliance*) presenta un informe enfocado a destacar las amenazas relacionadas con la naturaleza compartida y bajo demanda del *Cloud Computing*. A continuación se presentan las 9 amenazas más críticas que los expertos del sector han detectado (ordenadas por severidad) y se plantean contramedidas para poder minimizar los riesgos de dichas amenazas [8][9]:

1. **Violación de datos.** Se da cuando cualquier usuario malicioso o persona no autorizada entra en una red corporativa y roba datos confidenciales o sensibles. **Implicaciones:** La pérdida y la fuga de datos son dos graves amenazas para el *Cloud Computing*.

- Desafortunadamente, las contramedidas para mitigar una de ellas pueden exacerbar la otra. Es posible cifrar los datos para reducir el impacto por fuga pero si se pierde la clave de cifrado se perderán también los datos. Por otro lado, se pueden mantener copias *offline* para reducir el impacto de la pérdida de datos aunque, sin embargo, esto aumenta su exposición. **Soluciones:** (1) Aislar todas las máquinas virtuales y datos almacenados, (2) Borrar por completo sesiones de datos antes de entregar datos a nuevos usuarios para prevenir fuga de información y (3) Mantener copias *offline* de los datos.
- 2. Pérdida de Datos.** Los datos sensibles almacenados en el *cloud* pueden perderse de diversas maneras, incluyendo la eliminación accidental o corrupción de los datos almacenados. **Implicaciones:** Las normas de protección de datos, la destrucción de datos y la corrupción de datos personales se consideran formas de violación de datos y requieren sus correspondientes notificaciones. Normalmente, las políticas de cumplimiento requieren que las organizaciones mantengan registros de auditoría u otra documentación. Si una organización almacena estos datos en el *cloud*, la pérdida de los mismos podría poner en peligro el estado de cumplimiento de la organización. **Solución:** Usar herramientas DLP (*Data Loss Prevention*).
 - 3. Secuestro de cuenta o servicio.** Un intruso puede usar las credenciales robadas para secuestrar servicios de *Cloud Computing* logrando insertar datos falsos, desviar a los usuarios a sitios web abusivos, etc. **Implicaciones:** Se comprometen la confidencialidad, integridad y disponibilidad de los servicios ocasionando posibles problemas legales a los CSPs. **Soluciones:** (1) No compartir credenciales de cuentas entre empleados, (2) Aplicar técnicas de doble autenticación y (3) Realizar una buena definición de SLA.
 - 4. Interfaces y APIs inseguras.** Si las APIs utilizadas por los usuarios para comunicarse con los servicios en el *cloud* son débiles o no suficientemente seguras, un intento accidental o malicioso de violación de las mismas puede exponer los datos del *cloud* a múltiples amenazas. **Implicaciones:** Los clientes o usuarios de servicios en el *cloud* deben entender las implicaciones de seguridad asociadas al uso, gestión, orquestación y seguimiento de los servicios. Se pueden ver comprometidas la confidencialidad, integridad, disponibilidad y responsabilidad de los datos. **Soluciones:** (1) Evaluar la API antes de usarla y (2) Aplicar fuertes controles de acceso, autenticación y encriptar las transmisiones (por parte de los CSPs).
 - 5. Denegación de Servicio (DoS).** Se produce cuando se niega temporalmente el acceso a los servicios o datos almacenados en el *cloud* por parte de usuarios autorizados. Si un atacante genera miles de peticiones contra un servidor puede colapsarlo. **Implicaciones:** Las interrupciones de servicio pueden dar pie a reconsiderar si mover los datos críticos al *cloud* para reducir los costes de infraestructura realmente vale la pena. **Solución:** Utilizar sistemas de detección y prevención de intrusiones (IDS, IPS).
 - 6. Usuarios maliciosos con privilegios.** Un usuario con permisos de acceso a la red, al sistema o a los datos de una organización y que utiliza este acceso a los recursos de manera maliciosa, puede comprometer la confidencialidad, integridad y disponibilidad de los servicios. **Implicaciones:** Los sistemas cuya seguridad depende exclusivamente del CSP están en gran riesgo. Incluso si se implementa cifrado, si las claves no se mantienen con el cliente y sólo están disponibles cuando se usan los datos, el sistema sigue siendo vulnerable a usuarios maliciosos. **Solución:** Aplicar controles de nivel de acceso por usuario.
 - 7. Abuso de servicios en el cloud.** El *cloud* permite que organizaciones pequeñas tengan acceso a grandes cantidades de potencia de cálculo. Es complicado para una pequeña organización comprar y mantener decenas de miles de servidores, pero alquilarlos temporalmente a un CSP es mucho más asequible. Sin embargo, un cliente con propósitos maliciosos podría utilizar esta potencia computacional para descifrar claves en minutos, realizar ataques de DDoS (*Distributed Denial of Service*), etc. **Implicaciones:** Ataño a los CSPs y les plantea grandes retos. ¿Cómo detectar a las personas que abusan de su servicio? ¿Cómo definir “abuso”? ¿Cómo evitar que se repita? **Soluciones:** (1) Utilizar procesos de registro y validación de clientes antes de dar acceso al *cloud* y (2) Monitorizar pasivamente para asegurar que un usuario no afecta a otros.
 - 8. Insuficiencia por Diligencia.** Sin una comprensión completa del entorno del CSP (aplicaciones o servicios a implementar en el *cloud*, responsabilidades operacionales como respuesta a incidentes, cifrado y supervisión de seguridad) las organizaciones asumen niveles desconocidos de riesgo en formas que ni siquiera pueden comprender. **Implicaciones:** Los diseñadores y arquitectos de *cloud* deben estar familiarizados con las tecnologías empleadas para asegurar que los servicios trasladados a la nube no sean vulnerables. **Solución:** Establecer medidas de seguridad de datos, combinadas con transferencia del riesgo en forma de cobertura de seguro y aceptación de la asunción de riesgos por parte de los CSPs.
 - 9. Vulnerabilidades por Tecnología Compartida.** El aislamiento entre usuarios es complicado en una arquitectura *multi-tenant*. El CSP es responsable de entregar un servicio escalable a un cliente sin interferir con los sistemas de otros clientes. **Implicaciones:** El compromiso de un *hypervisor*, un componente de la plataforma compartida o una aplicación SaaS puede afectar a todo el entorno *cloud*. **Soluciones:** (1) Asegurar una fuerte compartimentación entre usuarios, (2) Aplicar mecanismos fuertes de autenticación y (3) Establecer SLA que contemple un remedio.
- A parte de conocer las amenazas que pueden afectar a la implementación de la solución *cloud* para el caso de uso *Smart Energy*, también resulta interesante disponer de un análisis de riesgos que especifique los aspectos de seguridad comprometidos. En [9] se realiza una clasificación de cada vulnerabilidad según riesgo y se asocian estas vulnerabilidades con los dominios CSA correspondientes al

documento “Security Guidance for Critical Areas on focus in Cloud Computing” [10] para aplicar las mejores prácticas en el momento de aportar seguridad a un diseño desde su concepción.

Por otro lado, puede ser de vital interés conocer las limitaciones de los problemas de seguridad a los que un cliente se expone para minimizar riesgos. Para ello, Gartner [11] propone siete ámbitos específicos sobre los cuales deben recoger información los clientes antes de seleccionar un CSP.

IV. PROBLEMAS DE SEGURIDAD EN EL CLOUD

En el entorno *cloud*, cada modelo de entrega de servicios tiene asociados sus problemas de seguridad. Es necesario conocer y analizar estos problemas para poder evaluarlos en el caso de uso *Smart Energy* de FINESCE según los modelos de entrega de servicios que finalmente se implementen.

A. Problemas de seguridad en IaaS

En un **modelo IaaS** primero de todo cabe destacar que no existen brechas de seguridad en el gestor de virtualización. El otro factor importante es la fiabilidad de los datos que se almacenan dentro del hardware del proveedor. Debido a la creciente virtualización de "todo", que el cliente propietario de los datos conserve el control final sobre éstos independientemente de su ubicación, se convierte en un aspecto de máximo interés. IaaS es propenso a diversos grados de problemas de seguridad basados en el modelo de despliegue a través del cual se está entregando el servicio. En el caso del *cloud híbrido* se deben tener en cuenta los aspectos de confiabilidad presentados en la Tabla I para gestionar correctamente la seguridad desde el diseño.

Tabla I
CONFIABILIDAD DEL MODELO IaaS SEGÚN SU DESPLIEGUE EN UN CLOUD HÍBRIDO [2]

	Gestión de la Infraestructura	Propietario de la infraestructura	Ubicación de la infraestructura	Acceso y consumo
Cloud Híbrido	Organización y terceros	Organización y terceros	En las instalaciones y fuera de las instalaciones	Confiable y desconfiable

B. Problemas de seguridad en PaaS

En el **modelo PaaS** el CSP puede otorgar algo de control a los desarrolladores de aplicaciones en la parte superior de la plataforma. Pero ningún tipo de seguridad por debajo del nivel de aplicación, como sistemas de prevención de intrusiones a nivel de red y host. El CSP debe poner especial atención en ofrecer fuertes garantías de que los datos permanecerán inaccesibles entre aplicaciones.

C. Problemas de seguridad en SaaS

En un **modelo SaaS** el cliente depende de que el proveedor aplique las medidas de seguridad adecuadas. Los datos de un usuario se almacenan en el CPD del proveedor de SaaS, junto con los datos de otros usuarios. Por otro lado, si el proveedor de SaaS está aprovechando un servicio de computación en el *cloud* público, los datos del usuario pueden almacenarse junto con los datos de otras aplicaciones SaaS no relacionadas. El CSP podría, además, replicar los datos en múltiples ubicaciones a través de diversos países a

efectos de proporcionar alta disponibilidad. Todo este funcionamiento del modelo SaaS provoca que el proveedor ponga todo su esfuerzo en asegurar al usuario el aislamiento de sus datos respecto a otros usuarios y en garantizar a dicho usuario que se están aplicando las medidas de seguridad adecuadas y que la aplicación estará disponible cuando sea solicitada. La Tabla II basada en [4] recoge los problemas de seguridad asociados al modelo SaaS mostrando una breve definición del entorno al que afecta cada problema y posibles soluciones a aplicar.

V. LA SEGURIDAD EN EL CLOUD DE FINESCE

El proyecto FINESCE forma parte de la segunda fase del proyecto FIWARE (*Future Internet Core Platform*) dentro del marco de financiación de FI-PPP (*Future Internet - Private Public Partnership*) del Séptimo Programa Marco de la Comisión Europea. FIWARE tiene como objetivo principal contribuir al desarrollo de una infraestructura abierta basada en las TIC utilizada para desarrollar nuevas soluciones y aplicaciones en todos los campos del Internet del Futuro, y FINESCE representa el caso de uso centrado en la aplicación de estas tecnologías para el sector energético. Entre otras actividades, FINESCE propone crear una infraestructura basada en un entorno de **cloud híbrido** como se muestra en la Fig. 4 para dar solución al caso de uso *Smart Energy*.

Se escoge un método de despliegue de *cloud híbrido* por las siguientes razones:

- Los datos sensibles como información crítica de la eléctrica se almacenan en el *cloud* privado para asegurar su confidencialidad.
- El *cloud* público se utilizará para almacenar datos no sensibles como históricos de medición de *smart meters*, datos de carga de vehículo eléctrico, etc., cuando los FIDEVs¹ (*FINESCE Devices*) estén llegando a su límite de almacenamiento.
- Existen aplicaciones susceptibles a latencias. El *cloud* privado evita las latencias imprecisas e incontrolables introducidas por Internet.

En la Fig.4 se puede apreciar que existen diversas zonas o regiones interconectadas a través de Internet. En zonas locales como Barcelona e Irlanda, se despliega un sistema de *cloud* privado mientras que el FIWARE LAB, plataforma de pruebas de FIWARE, representa la parte de *cloud* público. En cada región se sitúan dispositivos FIDEV que actúan como subestación virtual recogiendo datos de dispositivos conectados a la red eléctrica (*Smart Metering*, puntos de recarga de vehículo eléctrico, etc.). El mantenimiento de réplicas de información a través de los FIDEVs situados en los *clouds* privados de Barcelona e Irlanda y, también hacia el *cloud* público de FIWARE, pretende proporcionar una solución robusta de almacenamiento distribuido aprovechando las ventajas del *Cloud Computing* definidas en este documento.

¹ FIDEV. Es el dispositivo IDEV del proyecto INTEGRIS adaptado a las funcionalidades del Caso de uso *Smart Energy*.

Tabla II
PROBLEMAS DE SEGURIDAD EN ENTORNOS SAAS

Problema de Seguridad	Definición del entorno al que afecta el problema	Soluciones o Recomendaciones
<i>Seguridad de los datos</i>	<ul style="list-style-type: none"> Datos de usuario fuera de la empresa CSP controles adicionales para garantizar seguridad y evitar incumplimientos por vulnerabilidades aplicación o empleados maliciosos 	<ul style="list-style-type: none"> Fuertes técnicas de cifrado de datos y autorización Administradores sin acceso a instancias ni SO de cliente Registrar y auditar accesos rutinariamente
<i>Seguridad de red</i>	<ul style="list-style-type: none"> Datos sensibles de usuario procesados por aplicación y almacenados en el CSP 	<ul style="list-style-type: none"> Asegurar flujo de datos a través de la red para evitar fuga de información sensible Protección contra ataques <i>Man-in-the-middle</i>, suplantación IP, escaneo de puertos, detección de paquetes, etc. Técnicas de cifrado de tráfico de red SSL y TLS
<i>Ubicación de los datos</i>	<ul style="list-style-type: none"> Incertidumbre por parte del cliente de la ubicación de los datos almacenados Leyes de cumplimiento y privacidad de datos pueden variar entre países Jurisdicción de los datos cuando se produce una investigación 	<ul style="list-style-type: none"> El usuario debe asegurarse de cómo se aplican las leyes
<i>Integridad de los datos</i>	<ul style="list-style-type: none"> Comúnmente transacciones ACID para asegurar integridad de datos (Atomicidad, Coherencia, Aislamiento (<i>Isolation</i>) y Durabilidad) En sistemas distribuidos, mantener correcto manejo de datos y a prueba de fallos Entornos SOA, utilización SOAP y REST Utilización de HTTP que no permite transacciones o entrega garantizada 	<ul style="list-style-type: none"> Gestión centralizada de transacciones Implementar mecanismos de entrega garantizada a nivel de API WS-Transaction y WS-Reliability para integridad
<i>Segregación de datos</i>	<ul style="list-style-type: none"> Multi-tenancy, datos de varios usuarios en la misma ubicación Problemas de intrusión entre usuarios 	<ul style="list-style-type: none"> Garantía de mantenimiento de límites por parte del CSP tanto a nivel físico y de aplicación
<i>Acceso a datos</i>	<ul style="list-style-type: none"> En relación con las políticas de seguridad proporcionadas a los usuarios al acceder a los datos 	<ul style="list-style-type: none"> Cliente debe establecer políticas de seguridad de acceso a datos Asegurarse del cumplimiento de estas políticas por parte del CSP para evitar intrusiones no autorizadas CSP debe garantizar límites entre tenants
<i>Autenticación y autorización</i>	<ul style="list-style-type: none"> Comúnmente utilización servidores LDAP para acceso en grandes empresas y <i>Active Directory</i> en PyMEs Software gestión de usuarios alojado fuera de la empresa, credenciales usuarios almacenadas en bases de datos en CSP 	<ul style="list-style-type: none"> Cliente debe recordar eliminar/desactivar o crear/activar cuentas de empleados que dejan la empresa o empleados nuevos Si es necesario por seguridad, CSP puede delegar autenticación a LDAP/AD de la empresa
<i>Confidencialidad de datos</i>	<ul style="list-style-type: none"> Intercambio o almacenamiento de datos de servidores remotos pertenecientes o gestionados por terceros y accesibles a través de Internet u otras conexiones 	<ul style="list-style-type: none"> Establecer política de seguridad y SLA con CSP adaptada a los requisitos de confidencialidad y privacidad del usuario Mantener conocimiento de: (1) Derechos de privacidad aplicables según los datos entregados al CSP, (2) Obligaciones del CSP respecto privacidad y confidencialidad según ubicación datos, (3) Legalidad asociada a los datos según su ubicación
<i>Seguridad Aplicación Web</i>	<ul style="list-style-type: none"> Modificación de software de aplicación SaaS transparente para el usuario Si el software no está correctamente programado, los datos detrás de la aplicación SaaS y la propia aplicación estarán en riesgo 	<ul style="list-style-type: none"> Comprobar que la aplicación SaaS no es susceptible a las vulnerabilidades más relevantes señaladas en el Top 10 del OWASP Project [12]
<i>Violación de datos</i>	<ul style="list-style-type: none"> Datos sensibles de clientes almacenados en <i>cloud</i> 	<ul style="list-style-type: none"> Prohibir el acceso directo de los empleados del CSP a las bases de datos Controlar y monitorizar los accesos a cualquier parte del entorno <i>cloud</i> para evitar filtraciones de información sensible
<i>Virtualización</i>	<ul style="list-style-type: none"> Se presupone que diferentes instancias en una misma máquina virtual están aisladas entre ellas y de las tareas de virtualización 	<ul style="list-style-type: none"> Asegurar aislamiento Emplear VMs a nivel <i>root</i>, sin privilegios que permitan al huésped acceder al sistema anfitrión
<i>Disponibilidad</i>	<ul style="list-style-type: none"> Se presupone que aplicación SaaS desarrollada en arquitectura de varios niveles, con carga balanceada de instancias ejecutadas en varios servidores 	<ul style="list-style-type: none"> Aplicación SaaS desarrollada con resistencia a fallos de HW/SW y ataques DoS Disponer de un plan de acción para la continuidad de negocio y recuperación de desastres
<i>Copias de seguridad</i>	<ul style="list-style-type: none"> Se presupone que el CSP realiza copias regulares de los datos sensibles de sus clientes para facilitar rápida recuperación ante desastres 	<ul style="list-style-type: none"> Utilizar esquemas de cifrado robustos para proteger las copias de seguridad y evitar fugas de información
<i>Gestión de Identidad (IdM)</i>	<ul style="list-style-type: none"> Aplicaciones SaaS disponen de sistema que controla el acceso a los recursos mediante la colocación de restricciones sobre las identidades establecidas 	<ul style="list-style-type: none"> Mantener un sistema de gestión de identidades robusto Existen tres perspectivas a tener en cuenta en la aplicación de IdM: identidad pura, acceso de usuarios (log-on) y servicio

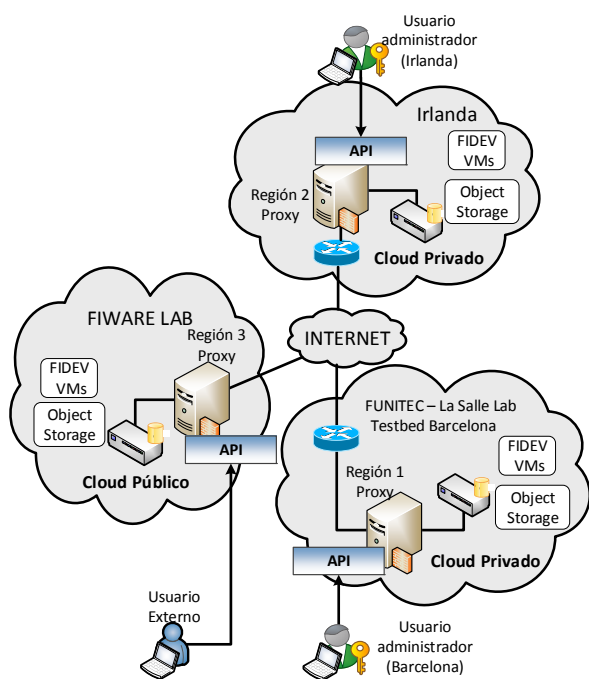


Fig. 4. Topología caso Smart Energy, proyecto FINESCE

Los FIDEVs incorporan un despliegue de *OpenStack* con funcionalidades de *Object Storage* para proporcionar el almacenaje de información y también las APIs necesarias para intercomunicarse con ellos. Estas APIs están basadas en *Generic Enablers (GEs)* definidos en el proyecto FIWARE y concretamente se utilizan las siguientes:

- En el *cloud* público se usa el *Object Storage GE* [13] para almacenaje y el *KeyRock GE* [14] como sistema de gestión de identidades que permite crear los usuarios y asociar autorizaciones de acceso a los FIDEVs para diferenciar los siguientes usuarios: (1) *Administradores*, empleados de la eléctrica con acceso a la gestión de toda la plataforma, (2) *Clientes de la eléctrica*, con acceso a sus propios datos a través del *cloud* público y (3) *Externos*, terceros con acceso únicamente a datos muy restringidos en el *cloud* público como históricos de consumo. Ambos GEs se consumen como SaaS de la plataforma FIWARE LAB.
- En el *cloud* privado se han desplegado instancias locales de los GEs anteriores según las especificaciones proporcionadas por FIWARE. De esta forma el resultado local está basado en un sistema de almacenamiento de objetos basados en SWIFT [15] y gestores de identidades basados en *KeyStone* [16], ambos módulos de la arquitectura *OpenStack*. En este *cloud* privado se consumen recursos en modo PaaS.

Definida la arquitectura de *cloud híbrido* para el caso de uso *Smart Energy*, es posible analizar los diversos componentes para ver si se ajustan a los requerimientos de seguridad básicos de un *cloud híbrido*.

A. Características de seguridad

Si se retoma la Fig. 3, los requerimientos de seguridad básicos para un *cloud híbrido* deben cumplir con las características de *Identificación* y *Autenticación*,

Autorización, *Confidencialidad* e *Integridad* para entornos SaaS e *Integridad* para entornos PaaS. Se cumplen todas las características básicas de seguridad de la siguiente manera:

- **Integridad.** Tanto *Object Storage* como SWIFT se encargan de almacenar los datos de manera íntegra.
- **Identificación, Autenticación y Autorización.** Cuando un usuario quiere realizar operaciones con datos contra el *cloud* privado (cargar, descargar, cifrar, descifrar) primero se autentica contra *KeyRock* que comprueba las credenciales y si ese usuario tiene autorización para la aplicación de almacenamiento que quiere usar, obtiene un *token* que se emplea para una segunda comprobación contra *KeyStone* de la infraestructura interna.
- **Autorización.** Las operaciones de transferencia de datos al *cloud* de FIWARE se realizan desde los FIDEVs locales, lo cual implica que el usuario esté autenticado previamente para realizar esta operación.
- **Confidencialidad.** Los datos se guardan confidencialmente mediante una clave definida por el usuario, de la cual se genera un hash mediante SHA-256 y ese hash se utiliza para cifrar los datos con AES-256.

Si se presta atención a la Tabla II, el sistema planteado pretende solucionar problemas relacionados con la *Seguridad*, *Integridad* y *Confidencialidad de los datos*, el *Acceso a los datos*, la *Autenticación* y *Autorización de usuarios*, la *Gestión de Identidades* y las *Copias de Seguridad*. El resto de problemas dependen más de los acuerdos mantenidos con el CSP mediante SLA y de la implementación de la propia infraestructura de *cloud* llevada a cabo por el CSP.

B. Problemas de Seguridad

Aunque exista un sistema de gestión de identidades que se preocupa de que usuarios no autorizados tengan dificultades para acceder a la plataforma y aunque se cifre la información cuando se almacena en los *Object Storage*, se han encontrado algunos puntos débiles en el sistema. El primer punto débil es la gestión de identidades mediante *KeyRock* que utiliza OAuth el cual es vulnerable a ataques *Cross-Site Request Forgery*, *Clickjacking* y *DoS*. El segundo punto débil se encuentra en la definición de las APIs de los GEs que tienen los siguientes problemas:

- CDMI (*Cloud Data Management Interface*) es la especificación que sigue la implementación de la API del *Object Storage GE*. Todas las interacciones que se realizan con el CDMI son mediante HTTP. Si un usuario malicioso es capaz de capturar el tráfico que se transmite hacia el *cloud* público podría obtener credenciales de usuario y también las claves con las cuales se cifra la información. Sin embargo, en el momento de transferir archivos al FIWARE LAB, como ya están cifrados, aunque se realice por HTTP no se podrán interpretar si no se dispone de la clave para descifrarlos. Las transacciones de datos entre el *cloud* público y privado suponen un problema de *Seguridad de Red* (Tabla II).
- Cuando un usuario se intenta autenticar contra el *Keystone* para realizar alguna operación, si el usuario es malicioso y no dispone de credenciales, puede utilizar ataques de fuerza bruta o de diccionario para acceder al sistema. La API no protege frente a intentos continuados.

- Para que un usuario pueda almacenar información en los objetos de almacenamiento, primero es necesario crear un contenedor que aloje sus archivos. La API del *Object Storage* tiene los siguientes problemas de funcionamiento respecto a los *containers* que causarían indisponibilidad del servicio: (1) Un usuario puede crear una cantidad de *containers* ilimitada, lo cual podría inutilizar el sistema para otros usuarios, (2) Aunque el tamaño de los archivos está limitado, no está limitado el tamaño de los *containers* que éste crea. Es posible que un solo usuario cargue archivos en su *container* hasta ocupar el espacio de almacenamiento total impidiendo que otros usuarios sean capaces de cargar sus archivos.
- Si se carga un archivo en el sistema con el mismo nombre que otro archivo cargado previamente, la API del *Object Storage* no notifica la colisión y sobrescribe el archivo antiguo con el nuevo. Por lo tanto, un usuario malintencionado que ha conseguido acceso al sistema o incluso, un usuario con garantías pero operando erróneamente, podría introducir sus archivos haciéndolos pasar por archivos que ya estaban almacenados. Si no se revisa el contenido de los archivos, no es posible apreciar que ha habido cambios en el almacenamiento.

VI. CONCLUSIONES

Las soluciones de *Cloud Computing* proporcionan una gran flexibilidad para los negocios actuales dado el amplio catálogo de soluciones (IaaS, PaaS, SaaS) que pueden implementarse interna o externamente en las organizaciones. Dada su eficiencia en costes al implementar modelos como SaaS, se ha convertido en una tendencia en crecimiento pero el hecho de que cada vez más usuarios dependan de un CSP pone a estos proveedores en el punto de mira de usuarios malintencionados. Cuando una organización piensa en trasladar información o aplicaciones al *cloud*, debe analizar detenidamente las amenazas y los riesgos a los cuáles se expone según el modelo de negocio que necesite. Sobre todo, es muy importante asegurar que la política de seguridad establecida en la organización se extienda hasta el *cloud* y para ello será necesario concretar un nivel de servicio estableciendo SLA entre el CSP y la organización. De esta manera la organización, operando como cliente de servicios *cloud*, podrá tener catalogados los riesgos a los que está expuesta.

Desde el punto de vista del *cloud híbrido* para el caso de uso *Smart Energy*, se ha comprobado la necesidad de someter la solución de *cloud* escogida a diversos análisis o auditorías para detectar amenazas y posibles riesgos, y tener la opción de contrarrestarlos antes de poner definitivamente el sistema en producción. Además, la dependencia del *cloud* público de FIWARE demuestra la importancia de conocer transparentemente cómo implementa las soluciones de *cloud* el proveedor para enfrentarse a problemas de *Ubicación, Segregación, Violación y Disponibilidad de los datos, Acceso a datos entre tenants, Virtualización y Seguridad de Aplicaciones Web*, los cuales dependen más de los acuerdos mantenidos con el CSP mediante SLA y de la implementación de la infraestructura física de *cloud* del CSP que no del sistema planteado. Al fin y al cabo, los servicios finales que una organización quiera proporcionar estarán condicionados a nivel de seguridad por el CSP.

Una vez analizada la plataforma y extraídas las vulnerabilidades que podrían ser explotadas por usuarios malintencionados, el principal objetivo de este proyecto es seguir mejorando el nivel de seguridad estudiando y probando soluciones que minimicen los riesgos asociados a estas vulnerabilidades.

AGRADECIMIENTOS

Este trabajo se está llevando a cabo gracias a la financiación de FINESCE, proyecto del séptimo programa marco (FP7) de la Comisión Europea bajo la call FLICT-2011, Grant number 604677. Los autores también agradecen a l'Agència de Gestió d'Ajuts Universitaris i de Recerca por su soporte al Grupo de Investigación en *Internet Technologies & Storage* (2014-SGR-589) y a La Salle – Universitat Ramon Llull.

REFERENCIAS

- [1] Mell JP., Grance T., "The NIST Definition of Cloud Computing", Version 15, National Institute of Standards and Technology, October 7, 2009. Available: <http://csrc.nist.gov/groups/SNS/cloud-computing>
- [2] A.B. Fernandes D., F.B. Soares L., V Gomes J., M. Freire M., R.M. Inacio P., "Security issues in Cloud Environment: A survey", *International Journal of Information Security*, 2013.
- [3] Bashir F., Haider S., "Security Threats in Cloud Computing ", 6th International Conference on Internet Technology and Secured Transactions, 11-14 December 2011.
- [4] Subashini S., Kavitha V., "A survey on security issues in service delivery models of cloud computing", *J Network Comput Appl* (2010), doi:10.1016/j.jnca.2010.07.006
- [5] Wei L., Zhu H., Cao Z., Dong X., Jia W., Chen Y., V. Vasilakos A., "Security and privacy for storage and computation in cloud computing", *Information Sciences*, Vol. 258, Pages(s) 371-386, 2014.
- [6] Dillon T., Wu C., Chang E., "Cloud Computing: Issues and Challenges", 24th IEEE International Conference on Advanced Information Networking and Applications, 2010
- [7] Amin A., Irfan M., Talib R., Sarwar U., "Security Issues in SaaS Delivery Model of Cloud Computing", *International Journal of Computer Science and Mobile Computing*, Vol. 3, Issue. 3, Page(s) 15-21, 2014
- [8] Venkata S., Padmapriya S., "A Survey on Cloud Computing Security Threats and Vulnerabilities", *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 2, issue 1, January 2014.
- [9] Cloud Security Alliance (2013) "Notorious Nine: Cloud Computing Top Threats". Link: <https://cloudsecurityalliance.org/download/the-notorious-nine-cloud-computing-top-threats-in-2013/>
- [10] Cloud Security Alliance (2011) "Security guidance for critical areas of focus in Cloud Computing V3.0". Link: <https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf>
- [11] Gartner (2008) "Assessing the Security Risks of Cloud Computing". Link: <https://www.gartner.com/doc/685308>
- [12] OWASP (2013) "Top Ten Project". Link: https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project#tab=OWASP_Top_10_for_2013
- [13] FIWARE Catalogue, "Object Storage GE". Link: <http://catalogue.fiware.org/enablers/object-storage-ge-fiware-implementation>
- [14] FIWARE Catalogue, "KeyRock". Link: <http://catalogue.fiware.org/enablers/identity-management-keyrock>
- [15] OpenStack, "SWIFT". Link: <https://wiki.openstack.org/wiki/Swift>
- [16] OpenStack, "Keystone". Link: <https://wiki.openstack.org/wiki/Keystone>

Two Approaches to Obtain Anonymity in Transferable Electronic Ticketing Schemes

Macià Mut Puigserver*, Jordi Castellà Roca†, M. Magdalena Payeras Capellà*, Julio Bondia Barceló†

* Dpt. Ciències Matemàtiques i Informàtica

Universitat de les Illes Balears, Carretera de Valldemossa, Km. 7,5, 07122, Palma, Spain

{macia.mut, mpayeras}@uib.cat

†Dpt. Enginyeria Informàtica i Matemàtiques, UNESCO Chair in Data Privacy,

Universitat Rovira i Virgili, Av. Països Catalans 26, E-43007 Tarragona, Spain

{jordi.castella, julio.bondia}@urv.cat

Resumen—Electronic tickets demonstrate, without the use of paper, the possession of an authorization or access to a service. In this scenario, some security requirements must be accomplished. Moreover, some services should guarantee the anonymity of the users in the system. In addition to these requirements, the transferability of a ticket from one user to another (without involving a third party) is very useful but also generates other issues to be solved in terms of security, as several attacks could be performed. In transferable tickets, the users must be anonymous even when the ticket is transferred, adding a lot of complexity to the system. In this article we present two proposals of an electronic ticketing system with anonymity and transferability, based on the use of RSA signatures or group signatures, respectively, thus presenting different levels of complexity and anonymity.

Palabras Clave—e-ticketing, e-commerce, transferability, privacy, security.

I. INTRODUCTION

The presence of information technologies is increasing progressively in our society. Users can access online services regardless of the place and time. For example, they can purchase a movie or theater ticket online. Nonetheless, in some cases, they have to print the ticket to have access to the service. In other words, the process is not completely electronic because a printed ticket is required. Thanks to the introduction of smartphones, all the processes can be performed electronically. These devices offer a good computational power, high storage capacity, and also different communication technologies, such as Near Field Communication (NFC). All these features can be available in a small device that allows mobility and flexibility, and which makes the system perfectly suitable for management, e-ticketing and e-payment schemes [5]. Electronic tickets can be defined as a representation of the owner's rights to act as a user of a service, thus preserving the same requirements than the ones offered in paper format. We would like to emphasize on the following properties: anonymity and transferability.

In the same way as paper tickets, electronic tickets have different properties according to the services where they are used. These services can be classified according to the anonymity offered. For instance, a plane e-ticket is non-anonymous: the identity of the passenger is a fixed parameter that is part of the e-ticket. In the case of e-tickets with revocable anonymity, the beneficiary can use the ticket as

the owner of it but without any need for identification. This modality helps to avoid fraud related to the reuse of e-tickets. E-tickets with non-revocable anonymity are not linked to a user at all. The user who owns this e-ticket is the one who can use this service. In transferable tickets several kinds of anonymity can be achieved, that is, weak or full depending on the features of the proposals.

Regarding the transferability property, there are several e-ticket sales and distribution companies that allow the e-ticket transfer^{1 2}. Nonetheless, the typical transfers of e-tickets are performed through a central service and are non-anonymous. We would like to transfer an electronic ticket in the same way that we can transfer a paper ticket, i.e. anonymously and without the participation of a central service.

We can find new systems by using cryptographic techniques that enable the online e-ticket issue and verification [4], [6], [8], [10], [11]. In [12], a recent implementation of electronic tickets over mobile devices with NFC technology has been performed.

However, users can not transfer their e-tickets to other users in [4], [11] and [12]. The transferability defined in [6] is carried out between transports systems. Users can use several transport systems (metro, bus, etc..) with the same e-ticket but they cannot transfer their e-tickets between them. Finally, the transferability provided in [8] and [10] does not allow the resolution of the conflict. Users can transfer the same e-ticket to several users (overspending) and they can prevent from being identified.

A. Contribution

A complete survey in this field can be found in [9]. The previous proposals, which have been already analyzed, use two different approaches to provide transferability to e-tickets: i) use a central service that synchronizes the transfer of e-tickets between users and does not allow revocable anonymity; ii) use a basic method of transfer without a complete conflict resolution system. In this case, it is not possible to identify the author of the overspending.

Thus, the main goal of our contribution is to preserve the security properties of (a) revocable anonymity, where user's

¹<http://www.ticketmaster.com/transfer>

²<https://www.e-ticket.lu>

identity could be only revealed in case of misbehaviour, (b) transferability, where electronic tickets can be transferred as a resale or loan without the collaboration of a central service, and (c) short-term linkability, where the user can easily demonstrate that she is the same user when receiving the ticket and later when transferring it. This can be achieved by using different approaches. In [1], we presented a proposal that uses short-term linkability with group signatures. We now present two approaches and we compare both of them in relation to security and anonymity. Both approaches, could allow to deploy transferable electronic ticketing in real scenarios, by fulfilling these security requirements.

II. BACKGROUND

In the first approach, we use digital signatures based on RSA. The second approach is based on the short group signature (BBS) scheme [3] in order to verify that a user is a correct member of a certain group of users. In this section, we introduce the main definitions related to the BBS signature, both the group signatures scheme and the Zero-Knowledge Proof (ZKP) of the group signatures for a better understanding of the first approach (please see [1] for further details). The background is not necessary in the case of RSA.

A. Group signatures scheme

$KeyGen_G$, $Sign_G$, $Verify_G$ and $Open_G$ are constructed from the same BBS scheme [3]. Both $SignLinkable_G$ and $VerifyLinkable_G$ have also been constructed in [7]. Consider bilinear groups G_1 and G_2 with respective generators g_1 and g_2 . The scheme uses a bilinear map $e : G_1 \times G_2 \rightarrow G_T$ and a hash function $H : \{0, 1\}^* \rightarrow \mathbb{Z}_p^*$. The public values are $g_1, u, v, h \in G_1$ and $g_2, w \in G_2$. Here $w = g_2^\gamma$ for some secret $\gamma \in \mathbb{Z}_p$.

The functions are:

- $KeyGen_G(n)$. This algorithm takes a parameter n as input, which is the number of members of the group.
- $Sign_G(gpk, gsk[i], M)$. Given a group public key $gpk = (g_1, g_2, h, u, v, w)$, a private user's key $gsk[i] = (A_i, x_i)$ and a message $M \in \{0, 1\}^*$, compute and output a signature of knowledge $\sigma = (T_1, T_2, T_3, c, s_\alpha, s_\beta, s_x, s_{\delta_1}, s_{\delta_2})$.
- $Verify_G(gpk, M, \sigma)$. Given a group public key $gpk = (g_1, g_2, h, u, v, w)$, a message M and a group signature $\sigma = (T_1, T_2, T_3, c, s_\alpha, s_\beta, s_x, s_{\delta_1}, s_{\delta_2})$, verify that σ is a valid signature of the message.
- $Open_G(gpk, gmsk, M, \sigma)$. Given a group public key $gpk = (g_1, g_2, h, u, v, w)$, the group master private key $gmsk = (\xi_1, \xi_2)$, a message M and a signature $\sigma = (T_1, T_2, T_3, c, s_\alpha, s_\beta, s_x, s_{\delta_1}, s_{\delta_2})$, it proceeds as follows. First, recover the user's A by performing $A \leftarrow T_3 / (T_1^{\xi_1} \cdot T_2^{\xi_2})$. If the elements $\{A_i\}$ of the user's private keys are given to the group manager, then she can look up the user index corresponding to the identity A recovered from the signature.
- $SignLinkable_G(gpk, gsk[i], M', \sigma, \alpha, \beta)$. Given a group public key gpk , a private user's key $gsk[i]$, a message M' , a previous signature σ , and the values α, β used for that signature, compute and output a signature σ' .

- $VerifyLinkable_G(\sigma, \sigma')$. This algorithm takes two signatures σ and σ' as input and outputs *true* or *false* depending on whether the signatures have been produced by the same signer's pseudonym: $(T_1 \stackrel{?}{=} T_1', T_2 \stackrel{?}{=} T_2', T_3 \stackrel{?}{=} T_3')$.

B. ZKP of the group signatures scheme

In our second proposal, linkable group signatures are used, as they enable to verify the internal message information as well as to verify that determined signatures related to the same event or e-ticket belong to the same *anonymous* user. However, the owner of the ticket must prove to the new owner that she is really the owner prior to the transfer of the e-ticket. Users demonstrate ticket ownership by means of a Zero Knowledge Proof (ZKP). The ZKP is based on the ticket signature and it does not require to compute a new signature. The ZKP comprises the following procedures: $ZKP_GCommit$, $ZKP_GResponse$ and $ZKP_GVerify$.

- $ZKP_GCommit(M^*)$. This procedure is performed by the user that wants to demonstrate (prover) to another user (verifier) that she is the right holder of the ticket. Given a public group key $gpk = (g_1, g_2, h, u, v, w)$, a group private key for the user $gsk[i] = (A_i, x_i)$ and a signed message $M^* = (M, \sigma)$, where $\sigma = (T_1, T_2, T_3, c, s_\alpha, s_\beta, s_x, s_{\delta_1}, s_{\delta_2})$, it generates the commitment $m' = (T_1, T_2, T_3, R'_1, R'_2, R'_3, R'_4, R'_5)$ as output.
- $ZKP_GResponse(m', c')$. Given a commitment m' , where $m' = (T_1, T_2, T_3, R'_1, R'_2, R'_3, R'_4, R'_5)$, and a challenge c' given by the verifier, the prover generates the response $s' = (s'_\alpha, s'_\beta, s'_x, s'_{\delta_1}, s'_{\delta_2})$.
- $ZKP_GVerify(m', c', s')$. Given a commitment m' , where $m' = (T_1, T_2, T_3, R'_1, R'_2, R'_3, R'_4, R'_5)$, a challenge c' given by the verifier, and the response $s' = (s'_\alpha, s'_\beta, s'_x, s'_{\delta_1}, s'_{\delta_2})$ provided by the prover, the verifier checks the correctness of the Zero Knowledge Proof.

III. DESCRIPTION OF THE SYSTEM

In this section, we describe our system proposal. First of all, the security requirements to achieve are briefly introduced in Section III-A. Then, the details of the first approach are presented in Section III-B while the details of the second approach are presented in Section III-C.

A. Requirements

We can classify e-ticket requirements into two categories [9]. On the one hand, we have security requirements, and on the other hand we have functional requirements for e-tickets. The main security requirements desired for our proposal are:

- **Authenticity:** the generated ticket is demonstrably genuine.
- **Non-repudiation:** the issuer of the ticket cannot deny its generation.
- **Integrity:** the ticket cannot be modified after its issue.
- **Revocable anonymity:** the ticket is anonymous, but if the user performs an unauthorized use, it can be then identified.
- **Non-overspending:** the ticket cannot be used more times than those established in the issue.

- Transferability: the ticket can be transferred to other users, without losing any of the requirements previously stated.

The functional requirements considered are the following:

- Validity time: the ticket can be valid until an established time.
- Online/Offline: ticket verification can require (or not) a connection to the Internet or a centralized system for checking.

B. Details of the protocol in the First Approach

This section describes the approach based on RSA signatures. One the one hand there are three main entities in the system: User (\mathcal{U}), Issuer (\mathcal{I}) and Service provider (\mathcal{P}). On the other hand, there are three more entities to manage the system: Registration Authority (RA) to assign a pseudonym to users, Certification Authority (CA) to create RSA public key certificates, and a Law Enforcement Authority (\mathcal{L}) to intervene in case of conflict. The protocol has seven phases:

1) *User credentials*: First, the user has to install the application of e-tickets on her smartphone. She must then register to make use of the various features that have been implemented in the e-ticket application. To do so, she has to contact with the registration authority (RA) to receive an alias ($Pseu_{\mathcal{U}}$). Afterwards, \mathcal{U} contacts the certification authority (CA) to get a public key certificate. Once this process has been completed, she can start using e-tickets to get access to the services offered. The registration protocol is carried out as follows:

1. \mathcal{U} :
 - a) generates her RSA key-pair (public, private);
 - b) creates a message with the user's information: $Credentials = (ID, CreditCardInformation, publicRSAkeyInfo, AdditionalInformation)$. The information in this message may vary depending on the final requirements of the implemented system;
 - c) sends $Credentials$ to RA ;
2. RA :
 - a) verifies the user's identity and the user's information and generates a new pseudonym ($Pseu_{\mathcal{U}}$) for this \mathcal{U} . Should verification fail, RA stops the protocol;
 - b) signs $\widehat{Pseu_{\mathcal{U}}} = Sign_{RA}(Pseu_{\mathcal{U}} || publicRSAkeyInfo)$; and
 - c) sends $Pseu_{\mathcal{U}}^* = (Pseu_{\mathcal{U}}, \widehat{Pseu_{\mathcal{U}}})$ to \mathcal{U} ;
3. \mathcal{U} :
 - a) sends $Pseu_{\mathcal{U}}^*$ to CA in order to obtain a public key certificate;
4. CA :
 - a) generates the user's certificate of her public RSA key: $cert_{\mathcal{U}}$; and
 - b) sends $cert_{\mathcal{U}}$ to \mathcal{U} .
The $cert_{\mathcal{U}}$ has the regular structure of an *X.509 Public Key Certificate*.

\mathcal{U} has now registered and she has a valid $cert_{\mathcal{U}}$ to use the e-ticket system.

2) *Ticket issue*: Once we have purchased it, the e-ticket exists only as a digital record. The buyer (\mathcal{U}) receives the equivalent to a real ticket in her mobile device which will be used to ensure access to a contracted service. Technically, the electronic ticket is an information structure with identifying data (unique to each ticket) and descriptive data (indicating the contracted service). To get an e-ticket, \mathcal{U} has to perform the following operations:

1. \mathcal{U} : sends a request to know what services are available: Req.
The Req message is a signed message with the user's certificate ($cert_{\mathcal{U}}$) attached to it;
 2. \mathcal{I} : verifies the user's credentials and sends the list of services to \mathcal{U} : List.
Should verification fail, \mathcal{I} stops the protocol;
 3. \mathcal{U} :
 - a) selects the service Sv from the List;
 - b) generates the ticket request $V^* = (V, \widehat{V})$, where $V = (Sv || n_{\alpha})$ and the signature request is $\widehat{V} = Sign_{\mathcal{U}}(V)$; where $Sign()_{\mathcal{U}}$ is a regular RSA signature operation using the \mathcal{U} 's private key; and
 - c) sends V^* to \mathcal{I} ;
 4. \mathcal{I} :
 - a) generates the ticket information: $T = (Sn, V^*, Ti, Tv, Tc)$, where the ticket includes Sn as the ticket serial number, V^* received from \mathcal{U} , Ti as the date of issue, Tv as the validity time, and Tc as the terms and conditions. The terms and conditions include information such as: ticket issuer, price, and whether the ticket is transferable or not.
 - b) signs the ticket: $T^* = (T, Sign_{\mathcal{I}}(T))$; where this signature is the standard RSA signature with the \mathcal{I} 's private key; and
 - c) sends the ticket T^* to \mathcal{U} ;
 5. \mathcal{U} :
 - a) verifies the signature of T^* .
 \mathcal{U} has now received a valid ticket from \mathcal{I} that gives the right to use the selected service.
- 3) *Ticket Transfer (first time)*: In this protocol, two users, \mathcal{U}_1 and \mathcal{U}_2 , follow the steps described below to transfer an electronic ticket. The owner, \mathcal{U}_1 , selects the desired ticket and generates a transfer request to a second user \mathcal{U}_2 . After the execution of this protocol, \mathcal{U}_2 has obtained the permission to use the ticket. The transfer works as follows:
1. \mathcal{U}_1 : sends the ticket (T^*) to \mathcal{U}_2 ;
 2. \mathcal{U}_2 :
 - a) verifies the information and signature of T^* ;
 - b) generates a random value $n_{\lambda_0} \xleftarrow{R} \mathbb{Z}_p$;
 - c) generates the first *transfer commitment* with the number of transferred times $k = 0$ and the agreed *price* for the transfer: $c_{\beta_0} = H(n_{\lambda_0} || k = 0 || price || flag_transfer)$;
 - d) signs it: $\widehat{c}_{\beta_0} = (Sign_{\mathcal{U}}(c_{\beta_0}))$; and
 - e) sends $c_{\beta_0}^* = (c_{\beta_0}, \widehat{c}_{\beta_0})$ to \mathcal{U}_1 ;
 3. \mathcal{U}_1 :
 - a) verifies the information and signature of c_{β_0} ;
 - b) generates a random value $n_{\beta_0} \xleftarrow{R} \mathbb{Z}_p$; and

- c) sends $C_{\beta_0} = \text{Sign}_{\mathcal{U}_1}(c_{\beta_0}^* || n_{\beta_0})$ to \mathcal{U}_2 ;
4. \mathcal{U}_2 :
- verifies that both signatures (C_{β_0} and $T.V^*$) have been performed by the same user (in this case, by \mathcal{U}_1).
- 4) *Ticket Transfer (following 'j' times):*
- \mathcal{U}_1 : sends $(T^*, C_{\beta_i}, c_{\beta_i}^*$ for $\{i = 0, \dots, (j-1)\}$) to \mathcal{U}_2 ;
 - \mathcal{U}_2 :
 - verifies the information inside the ticket and signature of T^* ;
 - verifies that signatures $T.V^*$, and C_{β_i} for $i = \{0 \dots (j-1)\}$ are chained properly;
 - generates a random value $n_{\lambda_j} \xleftarrow{R} \mathbb{Z}_p$;
 - generates the new *transfer commitment* with the number of transferred times $k = j$ and the agreed *price* for the transfer: $c_{\beta_j} = H(n_{\lambda_j} || k = j || \text{price} || \text{flag_transfer})$;
 - signs it: $\hat{c}_{\beta_j} = (\text{Sign}_{\mathcal{U}_1}(c_{\beta_j}))$; and
 - sends $c_{\beta_j}^* = (c_{\beta_j}, \hat{c}_{\beta_j})$ to \mathcal{U}_1 ;
 - \mathcal{U}_1 :
 - verifies the information and signature of $c_{\beta_j}^*$;
 - generates a random value $n_{\beta_j} \xleftarrow{R} \mathbb{Z}_p$; and
 - sends $C_{\beta_j} = \text{Sign}_{\mathcal{U}_1}(c_{\beta_j}^* || n_{\beta_j})$ to \mathcal{U}_2 ;
 - \mathcal{U}_2 :
 - verifies the signature C_{β_j} and verifies that this signature and the signature $\hat{c}_{\beta_{j-1}}$ have been performed by the same user (in this case, by \mathcal{U}_1).
- 5) *Ticket Verification (standard):* This protocol is used when no transfer has been performed since its issue. Here, \mathcal{U} shows the ticket to \mathcal{P} in order to be verified to receive the associated service. The protocol works as follows:
- \mathcal{U} sends the ticket T^* to \mathcal{P} ;
 - \mathcal{P} :
 - verifies the information and signature of T^* ; and
 - generates a random value $n_{\gamma} \xleftarrow{R} \mathbb{Z}_p$ and sends it back to \mathcal{U} ;
 - \mathcal{U}
 - generates $Y^* = (Y, \hat{Y})$ where $Y = (n_{\gamma} || T.Sn || \text{flag_spend_standard})$ and $\hat{Y} = \text{Sign}(Y)$; and
 - sends Y^* to \mathcal{P} ;
 - \mathcal{P} :
 - verifies the signature of \hat{Y} and checks that this signature and the signature $T.V$ are generated by the same user; and
 - stores $(T^*, Y^* = (Y, \hat{Y}))$ in \mathcal{P} 's centralized database.
- 6) *Ticket Verification (transferred):* The procedure to verify a transferred ticket is similar to the previous case. Additionally, \mathcal{P} has to check the *transfer commitment chain*. In the following protocol, the ticket has been transferred $(j+1)$ times:
- \mathcal{U} sends the ticket T^* to \mathcal{P} together with all C_{β_i} for $i = \{0 \dots j\}$ and ;
 - \mathcal{P} :
 - verifies the information and signature of T^* ; and
 - verifies the signature chain C_{β_i} for $i = \{0 \dots j\}$;
 - generates a random value $n_{\gamma} \xleftarrow{R} \mathbb{Z}_p$ and sends it back to \mathcal{U} ;

- \mathcal{U}
 - generates $Y^* = (Y, \hat{Y})$ where $Y = (n_{\gamma} || T.Sn || \text{flag_spend_transferred})$ and $\hat{Y} = \text{Sign}(Y)$; and
 - sends Y^* to \mathcal{P} ;
- \mathcal{P} :
 - verifies that both signatures (Y^* , $c_{\beta_j}^*$) have been generated by the same user; and
 - stores $(T^*, Y^*, C_{\beta_i} \forall i = \{0 \dots j\})$ in \mathcal{P} 's centralized database.

7) *Revocation of Anonymity:* The real identity of the user in this e-ticket system is hidden beneath a pseudonym ($\text{Pseu}_{\mathcal{U}}$). Moreover, the RA could allow any user the right to have more than one $\text{Pseu}_{\mathcal{U}}$ in order to make more difficult to trace the behaviour of the users. In any case, the system pretends to achieve what is usually called *weak anonymity* (it is unfeasible for an attacker to identify the spender of an e-ticket in a transaction) [2] in contrast to *strong anonymity* (it is unfeasible for an attacker to decide whether two transactions are made by the same user or not).

In this section, the steps to revoke the user anonymity are described:

- \mathcal{P} detects an overspending case: a single ticket has been spent more than once;
- Case 1: the ticket has been spent more than once using the same $\text{Pseu}_{\mathcal{U}}$
 - \mathcal{P} :
 - In this case, \mathcal{P} has found two Y^* with the signature of a same \mathcal{U} associated to a single ticket; then
 - sends (Y_1^*, Y_2^*) to a Law Enforcement Authority (\mathcal{L});
 - \mathcal{L} :
 - verifies the received information;
 - sends a request to RA to reveal the user identity of the \mathcal{U} linked to the pseudonym of the signatures in (Y_1^*, Y_2^*) ;
 - RA : reveals the user identity to \mathcal{L}
- Case 2: different users have spent the same e-ticket
 - \mathcal{P} :
 - In this case, \mathcal{P} has two Y^* associated to a single T^* , but they are signed by different users. So, the \mathcal{U} who has the right to spend it is the one who owns a longer signature chain: C_{β_i} for $i = \{0 \dots j\}$; then
 - The information associated to the ticket verification from both cases (two Y^* 's and T^*) is then sent to a Law Enforcement Authority (\mathcal{L});
 - \mathcal{L} :
 - verifies the information received;
 - sends a request to RA to reveal the user's identity of the \mathcal{U} linked to the shorter signature chain;
 - RA : reveals the user identity to \mathcal{L}

C. Details of the protocol in the Second Approach

This section describes the approach based on the short group signature (BBS) scheme [3]. On the one hand, there are four main entities in the system: User (\mathcal{U}), Issuer (\mathcal{I}), Service provider (\mathcal{P}) and Group Manager (\mathcal{M}_G), which only interacts in case of conflict. The protocol has six phases.

1) *Ticket issue*: In this protocol, \mathcal{U} receives a valid ticket from \mathcal{I} in order to be further used. The protocol works as follows:

1. \mathcal{I} generates and sends a random value $n_\alpha \xleftarrow{R} \mathbb{Z}_p$;
2. \mathcal{U} :
 - a) selects the service S_v ;
 - b) generates $V^* = (V, \widehat{V})$, where $V = (S_v || n_\alpha || \text{flag_issue})$ and $\widehat{V} = \text{Sign}_G(\text{gpk}, \text{gsk}[i], V)$ is the group signature;
 - c) sends V^* to \mathcal{I} ;
3. \mathcal{I} :
 - a) verifies the group signature: $\text{Verify}_G(\text{gpk}, V, \widehat{V})$;
 - b) generates the ticket information: $T = (S_n, V, T_i, T_v, T_c)$, where the ticket includes V received from \mathcal{U} , S_n as the ticket serial number, T_i as the date of issue, T_v as the validity time, and T_c as the terms and conditions;
 - c) signs the ticket: $T^* = \text{Sign}_{\mathcal{I}}(T)$, where this signature could be a standard RSA-like signature; and
 - d) sends the ticket T^* to \mathcal{U} ;
4. \mathcal{U} verifies the signature of T^* .

2) *Ticket transfer (first time)*: In this protocol, \mathcal{U}_1 transfers the original ticket to \mathcal{U}_2 by giving the permission to use it with a group signature which is linked to the commitment of the ticket V^* issued. The protocol works as follows:

1. \mathcal{U}_1 :
 - a) generates a commitment $m_{\beta_0} = \text{ZKPGCommit}(T^*)$; and
 - b) sends the commitment and the ticket $(m_{\beta_0} || T^*)$ to \mathcal{U}_2 ;
2. \mathcal{U}_2 :
 - a) verifies the information and signature of T^* : $\text{Verify}_{\mathcal{I}}(T^*)$;
 - b) verifies the group signature: $\text{Verify}_G(\text{gpk}, T.V, T.\widehat{V})$;
 - c) generates a random value $n_{\lambda_0} \xleftarrow{R} \mathbb{Z}_p$; and
 - d) generates the first *challenge* with the number of transferred times $k = 0$ and the *price* for the transfer agreed: $c_{\beta_0} = H(n_{\lambda_0} || k = 0 || \text{price} || \text{flag_transfer})$ and sends it to \mathcal{U}_1 ;
3. \mathcal{U}_1 :
 - a) generates the response $s_{\beta_0} = \text{ZKPGResponse}(m_{\beta_0}, c_{\beta_0})$;
 - b) generates a random value $n_{\beta_0} \xleftarrow{R} \mathbb{Z}_p$; and
 - c) sends $(s_{\beta_0} || n_{\beta_0})$ to \mathcal{U}_2 ;
4. \mathcal{U}_2 :
 - a) verifies the response: $\text{ZKPGVerify}(m_{\beta_0}, c_{\beta_0}, s_{\beta_0})$; and
 - b) generates $W_0^* = (W_0, \widehat{W}_0)$ where $W_0 = (n_{\beta_0} || T^* || \text{flag_transfer})$ and its group signature $\widehat{W}_0 = \text{Sign}_G(\text{gpk}, \text{gsk}[i], W_0)$ and sends it to \mathcal{U}_1 ;
5. \mathcal{U}_1 :
 - a) verifies the group signature: $\text{Verify}_G(\text{gpk}, W_0, \widehat{W}_0)$; and
 - b) generates $X_0^* = (X_0, \widehat{X}_0)$, where $X_0 = W_0^*$ and \widehat{X}_0 is a group signature which is linkable only to V^* :

$\widehat{X}_0 = \text{SignLinkable}_G(\text{gpk}, \text{gsk}[i], W_0^*, \widehat{V}, \alpha, \beta)$, and sends it to \mathcal{U}_2 ;

6. \mathcal{U}_2 :
 - a) verifies the group signature: $\text{Verify}_G(\text{gpk}, X_0, \widehat{X}_0)$ and
 - b) verifies that the two signatures have been performed by the same user: $\text{VerifyLinkable}_G(T.\widehat{V}, \widehat{X}_0)$. X_0^* works as a transfer agreement of the ticket of the user \mathcal{U}_1 to the user \mathcal{U}_2 .
- 3) *Ticket transfer (following 'k' times)*: In this protocol, \mathcal{U}_1 transfers the (already transferred) ticket X_{k-1} to \mathcal{U}_2 by giving the permission to use it with a group signature which is linked to the commitment of the ticket previously received. The protocol works as follows:
 1. \mathcal{U}_1 :
 - a) generates a commitment $m_{\beta_k} = \text{ZKPGCommit}(X_{k-1})$; and
 - b) sends the commitment and ticket $(m_{\beta_k} || X_{k-1})$ to \mathcal{U}_2 ;
 2. \mathcal{U}_2 :
 - a) verifies the information and signature of T^* : $\text{Verify}_{\mathcal{I}}(T^*)$, the group signature: $\text{Verify}_G(\text{gpk}, V, \widehat{V})$, and the linkability of the two group signatures of the beginning of the first transfer: $\text{VerifyLinkable}_G(\widehat{X}_0, T.\widehat{V})$;
 - b) for each transfer $\forall i \in [0, k)$, verifies the group signatures $\text{Verify}_G(\text{gpk}, X_i, \widehat{X}_i)$ and $\text{Verify}_G(\text{gpk}, W_i, \widehat{W}_i)$, and checks the linkability of $\text{VerifyLinkable}_G(\widehat{X}_i, \widehat{W}_{i-1})$;
 - c) generates a random value $n_{\lambda_k} \xleftarrow{R} \mathbb{Z}_p$; and
 - d) generates the first *challenge*: $c_{\beta_k} = H(n_{\lambda_k} || k || \text{price} || \text{flag_transfer})$ (with the price of the transfer agreed) and sends it to \mathcal{U}_1 ;
 3. \mathcal{U}_1 :
 - a) generates the response $s_{\beta_k} = \text{ZKPGResponse}(m_{\beta_k}, c_{\beta_k})$;
 - b) generates a random value $n_{\beta_k} \xleftarrow{R} \mathbb{Z}_p$; and
 - c) sends $(s_{\beta_0} || n_{\beta_0})$ to \mathcal{U}_2 ;
 4. \mathcal{U}_2 :
 - a) verifies the response: $\text{ZKPGVerify}(m_{\beta_k}, c_{\beta_k}, s_{\beta_k})$; and
 - b) generates $W_k^* = (W_k, \widehat{W}_k)$, where $W_k = (n_{\beta_k} || X_{k-1} || \text{flag_transfer})$ and its group signature is $\widehat{W}_k = \text{Sign}_G(\text{gpk}, \text{gsk}[i], W_k)$, and sends it to \mathcal{U}_1 ;
 5. \mathcal{U}_1 :
 - a) verifies the group signature: $\text{Verify}_G(\text{gpk}, W_k, \widehat{W}_k)$; and
 - b) generates $X_k^* = (X_k, \widehat{X}_k)$, where $X_k = W_k^*$ and \widehat{X}_k is a group signature which is linkable only to W_{k-1}^* : $\widehat{X}_k = \text{SignLinkable}_G(\text{gpk}, \text{gsk}[i], W_k^*, \widehat{W}_{k-1}, \alpha, \beta)$, and sends it to \mathcal{U}_2 ;
 6. \mathcal{U}_2 :
 - a) verifies the group signature: $\text{Verify}_G(\text{gpk}, X_k, \widehat{X}_k)$; and
 - b) verifies that the two signatures have been performed by the same user:

$VerifyLinkable_G(\widehat{X}_k, \widehat{W}_{k-1})$. X_k^* works as a transfer agreement of the ticket of the user \mathcal{U}_1 to the user \mathcal{U}_2 .

4) *Ticket verification (standard)*: This protocol is used when no transfer has been performed since its issue. Here, \mathcal{U} shows the ticket to \mathcal{P} in order to be verified to receive the associated service. The protocol works as follows:

1. \mathcal{U} sends the ticket T^* to \mathcal{P} ;
2. \mathcal{P} :
 - a) verifies the information and signature of T^* ; and
 - b) generates a random value $n_\gamma \xleftarrow{R} \mathbb{Z}_p$ and sends it back to \mathcal{U} ;
3. \mathcal{U} generates $Y^* = (Y, \widehat{Y})$, where $Y = (n_\gamma || T.Sn || flag_spend_standard)$ with a group signature which is linkable only to V^* :
 $\widehat{Y} = SignLinkable_G(gpk, gsk[i], Y, \widehat{V}, \alpha, \beta)$;
4. \mathcal{P} :
 - a) verifies the group signature: $Verify_G(gpk, Y, \widehat{Y})$ and verifies that the two signatures are generated by the same user:
 $VerifyLinkable_G(T, \widehat{V}, \widehat{Y})$; and
 - b) stores $T.Sn$ in \mathcal{P} 's centralized database.

5) *Ticket verification (transferred)*: This protocol is used when some transfer has been performed since its issue. Here, \mathcal{U} shows the ticket to \mathcal{P} in order to be verified and to receive the associated service. The protocol works as follows:

1. \mathcal{U} sends the transferred ticket X_k^* to \mathcal{P} ;
2. \mathcal{P} :
 - a) verifies the information and signature of X_k^* :
 $Verify_{\mathcal{I}}(T^*)$ and
 $Verify_G(gpk, X_k, \widehat{X}_k)$. The service provider \mathcal{P} can detect whether the ticket has been transferred or not depending on its content;
 - b) verifies that the two signatures which are included into the ticket have been generated by the same user:
 $VerifyLinkable_G(\widehat{X}_0, T, \widehat{V})$.
 - c) for all the transfers, $\forall i \in [0, k]$: verifies all the group signatures $Verify_G(gpk, X_i, \widehat{X}_i)$ and $Verify_G(gpk, W_i, \widehat{W}_i)$, and also the linkability of W_{i-1}^* : $VerifyLinkable_G(\widehat{X}_i, \widehat{W}_{i-1})$ where needed; and
 - d) generates a random $n_\gamma \xleftarrow{R} \mathbb{Z}_p$ and sends it back to \mathcal{U} ;
3. \mathcal{U} generates $Y^* = (Y, \widehat{Y})$, where $Y = (n_\gamma || T.Sn || flag_spend_transferred)$ and its linkable group signature to W_k^* as follows:
 $\widehat{Y} = SignLinkable_G(gpk, gsk[i], Y, \widehat{W}_k, \alpha, \beta)$;
4. \mathcal{P} :
 - a) verifies the group signature: $Verify_G(gpk, Y, \widehat{Y})$, and verifies that the two signatures are generated by the same user:
 $VerifyLinkable_G(\widehat{W}_k, \widehat{Y})$, where the ticket receiver then has to demonstrate that she is the same user both in the transfer and verification phases; and
 - b) stores $T.Sn$ in the \mathcal{P} 's centralized database.

6) *Revocation of anonymity*: To spend an e-ticket, any user has to do a new signature at the 3rd step of the ticket verification phase. In case of controversy (such as an e-

ticket overspending case), the group manager \mathcal{M}_G could take part in the resolution of the controversy and could revoke the anonymity of the signer that misbehaved by calling the $Open_G$ procedure.

IV. SECURITY AND TRANSFERABILITY OF THE SYSTEM

This security analysis is organized in four propositions that state the security features of the two approaches. The respective claims then discuss and provide evidence to support the propositions. The security arguments are presented for each approach if necessary. This discussion does not provide any demonstration of the security of the cryptographic primitives and does not pretend to be a formal analysis of the security of the protocol, but it substantiates the security properties of the protocol. Moreover, the results of this discussion will provide information to compare the benefits of each approach.

Proposition 1: Both approaches to obtain anonymity in transferable the e-ticketing system preserves authenticity, non-repudiation and integrity of the e-ticket.

Claim 1: It is computationally unfeasible to make a new fraudulent e-ticket.

Security Argument. First approach. A valid e-ticket has the form $T^* = Sign_{\mathcal{I}}(T)$. The \mathcal{I} generates the ticket information: $T = (Sn, V^*, Ti, Tv, Tc)$ and signs the ticket: $T^* = Sign_{\mathcal{I}}(T)$, where this signature is the standard RSA signature with the \mathcal{I} 's private key.

Security Argument. Second approach. A valid e-ticket has the form $T^* = (T, Sign_{\mathcal{I}}(T))$.

In both cases, the first step that \mathcal{P} does when an e-ticket is received is the verification of the signature. The Ticket Verification protocol will continue only if this verification is positive; otherwise, \mathcal{P} refuses \mathcal{U} 's request. Thus, making a new fraudulent valid e-ticket would be equivalent to breaking the signature scheme, which would be computationally unfeasible as we have supposed that \mathcal{I} uses a secure signature scheme.

Claim 2: The issuer \mathcal{I} cannot deny the emission of a valid e-ticket.

Security Argument. A valid e-ticket has \mathcal{I} 's signature, and the signature scheme used is secure. Consequently, the identity of the issuer is associated to the ticket i.e. the signature is a non-repudiation piece of evidence of her origin.

Claim 3: The content of the e-ticket cannot be modified.

Security Argument. If someone modifies the content of the ticket, a new \mathcal{I} 's signature then has to be generated over the modified content; otherwise, the e-ticket will not pass the verification i.e. it will not be valid. Again, if it is computationally unfeasible to forge \mathcal{I} 's signature, it is unfeasible to modify the content of the e-ticket.

Proposition 2: The e-ticketing systems described in §III-B and §III-C are anonymous. The offered service is revocable and anonymous.

Claim 4: The protocol to get an e-ticket is anonymous.

Security Argument. First Approach. After obtaining her credentials, \mathcal{U} contacts the \mathcal{I} in order to receive the e-ticket. \mathcal{U} attaches the credentials to the request. The request is signed

using a regular RSA signature using U 's private key. Since the credentials are issued by the credentials issuer, \mathcal{I} does not know the identity of U . However, if U uses its credentials in several requests, \mathcal{I} can link these requests. So the tickets are anonymous but linkable until U renews its credentials. For this reason, the anonymity achieved by this approach is a weak anonymity.

Security Argument. Second Approach. The user establishes a connection with the ticket issuer \mathcal{I} in order to receive the e-ticket. Additionally, the user does not use any personal authentication method to get the e-ticket. \mathcal{I} generates and sends the e-ticket to U if she accredits to be a member of the group of users by producing a valid group signature over a challenge sent by \mathcal{I} . The $Verify_G(V)$ procedure performed by \mathcal{I} cannot identify the user due to the group signatures properties. For this reason, the anonymity achieved by this approach is a strong anonymity.

Claim 5: An e-ticket has revocable anonymity.

Security Argument. First Approach. The e-ticket is generated by \mathcal{I} , who only knows the user's credentials as we have discussed in the previous claim. The only item inside the e-ticket that can be used to identify the user is the ticket request $V^* = (V, \hat{V})$, where $V = (Sv || n_\alpha)$ and the signature request is $\hat{V} = Sign_U(V)$. Since the proposal presents weak anonymity, the RA can reveal the user's identity by using this information included in the e-ticket if it is required by \mathcal{L} .

Security Argument. Second Approach. A valid e-ticket does not have any information related to the user's identity. The e-ticket is generated by \mathcal{I} , who does not know the user's identity as we have discussed in the previous claim. The only item inside the e-ticket that can identify the user is the group signature V , and only the group manager \mathcal{M}_G can reveal this information by performing $Open_G(gpk, gmsk, M, \sigma)$, because it is the only entity that knows $gmsk$. Therefore, e-tickets can be spent anonymously but the anonymity can be revoked by \mathcal{M}_G . \mathcal{M}_G plays the role of a trusted third party, thus it will only do that by law enforcement.

Claim 6: In spite of the anonymity of the e-ticket, a fake user cannot spend an e-ticket impersonating another user.

Security Argument. First Approach. In order to spend the e-ticket, the legitimate U has to prove the ownership of the e-ticket with the use of a pair of RSA signatures. The provider verifies the information and the signature of $T^* = Sign_{\mathcal{I}}(T)$. In the standard verification, the provider generates a random value and the user signs a new element containing the random number using her private RSA key. Finally, the provider checks whether the signature included in the e-ticket and the new one are generated by the same user. If this is not the case, the e-ticket will not be accepted. In the verification of transferred tickets, a similar procedure is used. See proposition 4.

Security Argument. Second Approach. In order to spend the e-ticket, the legitimate U has to prove the ownership of the e-ticket with the use of a linkable signature Y with the element V , placed inside the e-ticket. An illegitimate U cannot perform the $SignLinkable_G$ operation properly: the fraud will be detected because the $VerifyLinkable_G()$ operation performed in the ticket verification protocol will warn \mathcal{P} about this impersonation attack.

Proposition 3: The protocol controls overspending.

Claim 7: If the ticket is only validated by one \mathcal{P} , the verification can be offline and \mathcal{P} can control any overspending attempt.

Security Argument. In this case, the security argument is the same for both approaches. \mathcal{P} maintains a database with the serial numbers of the e-tickets already validated (i.e used). \mathcal{P} can check both the issuer's signature and whether the e-ticket has not been spent before by using the information stored in the database. So the provider does not need to contact any party during the validation of an e-ticket.

Claim 8: If the ticket is validated with several providers, all \mathcal{P} s must then be connected and must share a database of spent tickets.

Security Argument. In this case, the security argument is similar for both approaches. The set of providers maintains a shared database with the serial numbers of the e-tickets that have been already validated. The contents of this database are used by the providers to decide whether they accept and validate a new ticket. So the provider does not need a connection with the issuer during the verification of an e-ticket, but the set of providers must share a database instead. If a persistent connection is not available, the overspending can then be detected and the identity of the overspender can be revealed by the group manager through the $Open_G(gpk, gmsk, M, \sigma)$ procedure in the second approach, and the credentials issuer in the first approach. Thus, the verification can be offline.

Proposition 4: Users can transfer their e-ticket to other users by using of the proposed scheme. The transferability operation among users preserves the security properties no matter how many transfers of the e-ticket have been made.

Claim 9: A transferred e-ticket can guarantee authenticity, non-repudiation and integrity properties as a non-transferred e-ticket.

Security Argument. First Approach. For each transfer, a new transfer commitment is added to the ticket, including the number of transfers and the agreed price for the transfer, and it is signed by the user. Since a new RSA signature is added to the e-ticket, the properties of the signature keep authenticity, non-repudiation and integrity of the transferred e-ticket since the discussion of the proposition 1 is already valid.

Security Argument. Second Approach. During a transfer, the format of the e-ticket is not substantially altered. Only a new group signature of the new owner is added to the e-ticket so that the properties of the signature keep authenticity, non-repudiation and integrity of the transferred e-ticket, since the discussion of the proposition 1 is already valid.

Claim 10: A transferred e-ticket preserves the anonymity of its owner.

Security Argument. First Approach. During a transfer, the owner of the e-ticket signs the transfer commitment using her credentials. The receiver of the transferred e-ticket checks that the signature has been generated by the same user who obtained the ticket during the issue or a previous transfer. The identity of this signer is protected by the anonymous credential, so the weak anonymity is preserved during the transfer phase. The receiver of the ticket also signs a commitment in a similar way as in the issue of the ticket, so the anonymity of

a transferred ticket is the same anonymity of an issued ticket, in this case weak anonymity due to the possibility of linkage. *Security Argument. Second Approach.* During a transfer, the owner of the e-ticket proves its ownership with a ZKP operation in order not to disclose her identity. The receiver of the transferred e-ticket includes a new group signature in it, so her identity is similarly protected in the same way, as we see in the step 4 of the ticket verification protocol (transferred version).

Claim 11: A transferred ticket cannot be overspent.

Security Argument. First Approach. The overspending detection procedure described in proposition 3 is also valid for transferred e-tickets as well as the anonymity revocation which can be made through the credentials issuer because this property relies on the verification made by \mathcal{P} on the serial number of the e-ticket stored in the database. The transfer of any e-ticket does not change its serial number and the provider receives the transfer commitment chain that includes all the signatures, and it can easily identify the user that has overspent the ticket.

Security Argument. Second Approach. The overspending detection procedure described in proposition 3 is also valid for transferred e-tickets as well as the anonymity revocation which can be made using $Open_G(gpk, gmsk, M, \sigma)$, because this property relies on the verification made by \mathcal{P} on the serial number of the e-ticket stored in the database. The transfer of any e-ticket does not change its serial number.

V. COMPARISON OF THE APPROACHES AND CONCLUSIONS

We have presented two approaches to obtain anonymity in transferable electronic ticketing systems. These approaches guarantee the anonymity for their users even when the system allows the transfer of tickets between them through payment or loan. We have to take into account that the kind of attacks that can be performed in a transferable system are greater than those in a non-transferable system due to the possibility of attack by confabulated users.

The first approach provides anonymity using RSA signatures and a registration authority. A transfer commitment chain is generated to maintain anonymity as well as security during the transfers. Users are not identified if they use the system properly. Moreover, the system allows the revocation of the identity of fraudulent users. However, if the user uses her credentials several times, the tickets obtained can be linked. For these reason, the anonymity provided by this approach is called weak anonymity.

The second scheme proposed is fully anonymous because a group signature scheme has been used. It allows the issuer to verify that the user belongs to a valid group of users, yet cannot identify which one she is. If the user tries to commit fraud, the group manager can revoke her anonymity. To achieve transferability, we use a linkable group signature scheme. With this technique, group signatures from the users involved in the transfer operation are used in order to generate a ticket transfer agreement, which could be further used as evidence in case of conflict between the parties.

As future work, the main goal will be to develop and evaluate the performance of the protocols in a mobile platform, in order to check its feasibility. The performance of the system

will be a very important point because users have to be able to obtain, validate or transfer the ticket in a short time. For this reason, we present two approaches with different degrees of anonymity. In a future work, we will evaluate whether the second approach, more robust, can be efficiently used in current devices and whether the increased complexity of the second approach is justified in order to obtain a greater degree of anonymity.

ACKNOWLEDGEMENTS

This work was partially supported by the Spanish Government under CO-PRIVACY TIN2011-27076-C03-01, AccessTur TIN2014-54945-R, SmartGlacis TIN2014-57364-C2-R and MobileKey RTC-2014-2552-7 projects. Some of the authors are members of the UNESCO Chair in Data Privacy, yet the views expressed in this paper neither necessarily reflect the position of the UNESCO nor commit with that organization.

REFERENCIAS

- [1] A. Vives-Guasch and M. Payeras-Capellà and M. Mut Puigserver and Jordi Castellà-Roca and J. Ferrer-Gomila, "Anonymous and Transferable Electronic Ticketing Scheme", En Data Privacy Management and Autonomous Spontaneous Security, DPM 2013, LNCS 8247, Pages 100–113, Springer 2013.
- [2] Blazy, Olivier and Canard, Sébastien and Fuchsbaauer, Georg and Gouget, Aline and Sibert, Hervé and Traoré, Jacques, "Achieving Optimal Anonymity in Transferable E-Cash with a Judge", In Progress in Cryptology, AFRICACRYPT 2011, volume 6737 of Lecture Notes in Computer Science, pages 206-223. Springer Berlin Heidelberg, 2011.
- [3] D. Boneh, X. Boyen and H. Shacham, "Short group signatures", In CRYPTO, volume 3152 of LNCS, pages 41-55. Springer, 2004.
- [4] Y. Chen, C.-L. Chen, and J.-K. Jan, "A mobile ticket system based on personal trusted device". Wireless Personal Communications, 40(4):569-578, 2007.
- [5] S. Ghiron, S. Sposato, C. Medaglia, and A. Moroni. "Nfc ticketing: A prototype and usability test of an nfc-based virtual ticketing application". In Near Field Communication, NFC '09, p. 45-50. IEEE, 2009.
- [6] T. S. Heydt-Benjamin, H.-J. Chae, B. Defend, and K. Fu. "Privacy for public transportation". In 6th Workshop on Privacy Enhancing Technologies (PET 2006), LNCS 4258, p. 1-19. Springer, 2006.
- [7] A. P. Isern-Deya, Vives-Guasch, M. Mut-Puigserver, M. Payeras-Capella and J. Castella-Roca. "A secure automatic fare collection system for time-based or distance-based services with revocable anonymity for users". The Computer Journal, 2012.
- [8] O. Jorns, O. Jung, and G. Quirchmayr. "A privacy enhancing service architecture for ticket-based mobile applications". In Availability, Reliability and Security, pages 374-383, 2007. ARES 2007 - The International Dependability Conference. vol. 24.
- [9] M. Mut-Puigserver, M. M. Payeras-Capella, J.-L. Ferrer-Gomila, A. Vives-Guasch, and J. Castella-Roca. "A survey of electronic ticketing applied to transport. Computers and Security, 31(8):925-939, 2012.
- [10] D. Quercia and S. Hailes. "Motet: Mobile transactions using electronic tickets". In Security and Privacy for Emerging Areas in Communications Networks, Proceedings, pages 374-383, 2005. vol. 24.
- [11] A. Vives-Guasch, J. Castella-Roca, M. Payeras-Capella, and M. Mut. "An electronic and secure automatic fare collection system with revocable anonymity for users". In Advances in Mobile Computing and Multimedia (MoMM), 2010.
- [12] A. Vives-Guasch, M. M. Payeras-Capella, M. Mut-Puigserver, J. Castella-Roca, and J. L. Ferrer-Gomila. "A secure e-ticketing scheme for mobile devices with near field communication (nfc) that includes exculpability and reusability". IEICE Vol.E95-D No.1, 2012.

Comparación de modelos de procedimiento en el análisis forense digital

Jordi Blanco-Galdeano, Helena Taribó-Gómez, Juan Vera-del-Campo, Miguel Soriano
Departament d'Enginyeria Telemàtica,
Universitat Politècnica de Catalunya BarcelonaTECH.
Jordi Girona 1-3, 08034. Barcelona.
jordibg89@gmail.com, helenatari@gmail.com, juanvi@entel.upc.edu, soriano@entel.upc.edu

Resumen—El análisis forense digital da una respuesta legal a las acciones ilegales que de alguna manera involucren ordenadores u otros dispositivos digitales. Los modelos de procedimiento a seguir durante un análisis forense permiten trabajar de forma organizada durante todo el proceso, asegurar la cadena de custodia de las evidencias recogidas tras la incidencia y elaborar el informe resultado del análisis. Este trabajo demuestra la importancia de un procedimiento correcto a la hora de realizar este tipo de investigaciones forenses. Para ello, se muestran distintos modelos de procedimiento y se comparan entre sí con el fin de analizar sus aspectos más importantes.

Palabras Clave—digital forensics, análisis forense digital, comparativa

I. INTRODUCCIÓN

El análisis forense digital nos permite recabar pruebas que permitan demostrar delitos digitales tales como fraudes digitales, *cyber bullying*, *spam* o *phishing* entre otros. Pero su alcance va más allá de este tipo de crímenes. El rastro que todas las personas dejamos en los dispositivos informáticos hace que la investigación de, por ejemplo, ordenadores o teléfonos móviles, nos den valiosas pruebas de cualquier tipo de crimen aunque a priori no parezca relacionado con el mundo digital.

El proceso forense digital consiste principalmente en recoger las pruebas y analizarlas, siempre conservando su integridad para que puedan ser válidas en un juicio. La cadena de custodia es el proceso que va de la recolección de la prueba hasta su exposición en el juicio. En este proceso debe quedar claro en todo momento qué acciones se realizan con la prueba y qué personas intervienen. Todo esto debe estar debidamente documentado de forma que no haya ninguna ambigüedad y de esta manera la prueba tenga validez.

Existen muchos modelos de procedimiento forense como el de Reith et al. [1], el de Freiling et al. [2], el de Agarwal et al. [3], o el de Rogers et al. [4]. Sin embargo, no existe un modelo que haya adquirido un carácter general sobre cómo se realiza el proceso. Durante años, varios modelos de procedimiento han sido propuestos y ninguno se ha establecido como modelo generalizado a seguir. En parte, esto se debe a que los diferentes agentes que intervienen en el proceso son profesionales de campos diferentes que si bien deben estar coordinados, a veces no tienen que seguir un objetivo común. Por ejemplo, la forma en que un policía obtiene la prueba no debe interferir en cómo el experto del laboratorio trabaja. Los agentes involucrados en una investigación, aunque sí deben seguir unas normas generales, no tienen por qué estar sujetos al mismo modelo.

Un buen modelo de procedimiento debe contemplar las distintas fases, desde la recolección de las pruebas hasta su exposición final. Debe dejar claro las normas en el momento de la adquisición de las pruebas y en su análisis, pero sobre todo dejar claro cómo debe documentarse y prepararse de cara a un posible juicio. La documentación es una parte vital del proceso ya que servirá para dejar constancia de cada uno de los pasos realizados, de esta manera quedará absolutamente claro que no ha habido ninguna irregularidad en el proceso y todo se ha hecho de forma correcta. Además, será una de las principales fuentes de información de cara a un posible juicio, así que es muy necesario que el modelo de procedimiento recoja instrucciones precisas sobre cómo documentar cada paso así como las conclusiones finales.

En la sección II se introducen los conceptos básicos del análisis forense, así como las diferentes fases que aparecen como respuesta a un incidente. En la sección III se introducen varios modelos de análisis digital forense. Estos modelos son comparados en la sección IV. Este trabajo acaba con las conclusiones y las referencias que hemos utilizado.

II. ANÁLISIS FORENSE DIGITAL

En esta sección se definirán conceptos básicos sobre el análisis forense digital; la propia definición de análisis forense digital y la definición de prueba digital.

II-A. Conceptos básicos

Análisis forense digital: el análisis forense digital se basa en el uso de métodos científicos, con una base legal, para preservar, recolectar, validar, identificar, analizar, interpretar y presentar pruebas que provienen de una fuente digital. Tiene el objetivo de facilitar la reconstrucción de acontecimientos criminales o ayudar a anticipar acciones no autorizadas que puedan dañar nuestros sistemas.

Pruebas digitales: hay muchas definiciones de prueba digital, ésta es una de ellas. Según el Departamento de Justicia de los Estados Unidos de América, una prueba digital es una unidad de información, o un dato, de valor para una investigación que se guarda, recibe o envía mediante un dispositivo electrónico. Esta prueba se consigue cuando los datos o los dispositivos electrónicos se incautan y se protegen para su posterior análisis.

II-B. Procedimiento del estudio forense digital

Para gestionar un caso de la mejor forma posible, es importante seguir una serie de fases que se agrupan en un modelo de procedimiento del estudio forense digital. No hay un modelo

estándar sino que se han presentado diferentes propuestas a lo largo de más de 20 años, pero todos ellos remarcan la importancia de preservar las pruebas. Se ha escogido el modelo propuesto por Michael Donovan Köhn en [6] porque incluye la fase de la preparación, fase que debería ser relevante tanto por las empresas como para los particulares dado que los sistemas de seguridad tienen debilidades.

1. Documentación: La documentación no es una fase en sí misma pero se ha incluido como primer paso porque es muy importante mantener un registro de todo lo que se hace y se observa a lo largo de toda la investigación con el fin de mantener la cadena de custodia.
2. Preparación: La fase de preparación es importante para empresas y usuarios en general, y tiene que describir cómo actuar en caso de detectar una incidencia. El objetivo del plan de preparación es el de maximizar la recolección de pruebas del entorno del incidente a la vez que se intenta minimizar el daño recibido. Las políticas de empresa tienen que describir de forma clara cómo los empleados deben tratar con las pruebas con el fin de poder ser analizadas. El objetivo de estos documentos es garantizar la integridad durante la investigación cuando se manipulen las pruebas.
3. Incidencia: El primer paso de esta fase es detectar la incidencia, ya sea de manera automática o manual. Una incidencia es una acción llevada a cabo con el objetivo de comprometer la confidencialidad, disponibilidad e integridad de la información. Una vez detectada, se necesita confirmar la incidencia para determinar si se trata de falsos positivos. Una vez confirmada la incidencia, se deben avisar a los investigadores para que empiecen la investigación.
4. Respuesta a las incidencias: La fase de respuesta empieza con la llegada de los investigadores, que necesitan definir una estrategia a seguir para asegurarse de que se mantienen las cadenas de custodia y de evidencia, y minimizar el posible daño a potenciales pruebas digitales. Los investigadores, cuando llegan a la escena del crimen tienen que buscar y preservar todos los datos encontrados para, posteriormente, poder analizarlos. Todas las pruebas físicas, como discos duros, las guardan las autoridades que correspondan según el tipo de investigación.
5. Análisis forense digital: Este paso será explicado en profundidad en la sección III de este documento, mediante varios ejemplos.
6. Presentación: Finalmente, se presenta un informe escrito con la recopilación de los procesos llevados a cabo desde el momento de toma de posesión de los datos. Este informe tiene que incluir una explicación detallada de todo lo que se ha hecho y las conclusiones técnicas a las que se ha llegado. Independientemente de si los destinatarios de este informe son los abogados y directivos de una empresa o es el juez, el informe tiene que incluir descripciones dirigidas a personas sin conocimientos técnicos.

III. ANÁLISIS FORENSE DIGITAL: MODELOS DE PROCEDIMIENTO

Como se ha comentado anteriormente, muchas han sido las propuestas para un modelo de procedimiento a seguir durante una investigación forense. En esta sección vamos a ver algunos de ellos.

Para este análisis, se han escogido distintos tipos de modelo con diferentes números de pasos y centrados en diferentes conceptos, que sean recientes y estén enfocados de distinta manera para así poder analizar sus diferencias.

III-A. Modelo 1

Extended Abstract Digital Forensic Model with 2PasU del año 2014 [5]. Este modelo está basado en [1], haciendo especial hincapié en la preservación y seguridad de las pruebas. Este modelo se enfoca principalmente a la exposición del caso en un juicio ya que hace continuas menciones a esta finalidad durante los pasos del proceso. El modelo consta de siete fases: (1) preparación y planteamiento, (2) recolección, (3) examen, (4) análisis, (5) informe, (6) presentación, (7) archivar y devolver la prueba. El núcleo del estudio forense está en los puntos (2), (3), (4) y (5) mientras que en los otros puntos encontramos pasos importantes para completar el trabajo forense. Podemos ver el esquema de las fases en la Fig. 1. A continuación se exponen cada uno de los pasos nombrados anteriormente:

1. Preparación y planteamiento: la primera de las fases no es exactamente una parte del proceso forense, pero es un punto de partida para trabajar. Esta fase consiste en conseguir toda la información posible previa al estudio para diseñar una estrategia a seguir. Los siguientes pasos se verán condicionados a las decisiones tomadas en este punto.
2. Recolección: este paso trata, en primer lugar, de preservar la prueba física. Para esto, una vez recibida la prueba, se hacen las copias necesarias de la prueba digital para poder trabajar con las copias sin poner en riesgo la prueba original. A continuación, se puede iniciar la recaudación de datos a partir de las copias.
3. Examen: esta fase consiste en identificar los datos recolectados en la fase anterior. De este modo podremos decidir qué datos son importantes para hacer el análisis posteriormente. Esta fase de examen puede hacerse al mismo tiempo que la fase anterior pero para generalizar el modelo se separa en dos puntos distintos.
4. Análisis: en este punto se deben tener los datos relevantes listos para trabajar con ellos. Estos datos se analizarán para ver si pueden ser usados como pruebas en un juicio. Si los resultados en este punto no son del todo satisfactorios, es posible que se deba volver al punto 2 las veces que sean necesarias. De este modo, se tiene la opción de encontrar datos que no habían sido hallados o dar importancia a otros que habían pasado desapercibidos.
5. Informe: una vez se ha completado un análisis satisfactorio, es el momento de redactar las conclusiones obtenidas como experto. Estas conclusiones deben estar redactadas de forma que sean entendibles por un juez. El

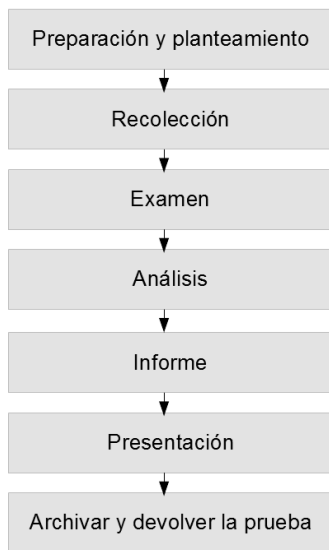


Figura 1. Esquema del modelo 1

informe también debe explicar todos los pasos realizados y demostrar que la prueba no ha sido contaminada.

6. Presentación: esta fase consiste en exponer las conclusiones de forma que quede claro que el trabajo que se ha hecho es correcto y convencer que la postura que se defiende en el juicio es la correcta.
7. Archivar y devolver la prueba: esta última fase consiste en devolver la prueba a quien le pertenezca, manteniendo siempre toda la protección sobre la prueba para que pueda ser usada posteriormente por cualquiera que lo necesite. En esta fase también se contempla la posible destrucción de los datos para preservar los derechos de las personas investigadas.

III-B. Modelo 2

Integrated Digital Forensic Process Model (IDFPM) del año 2012 [6]. Este modelo está realizado a partir de varios modelos anteriores intentando recoger lo mejor de cada uno de ellos. Este modelo separa cada parte del proceso de modo que queden muchos pasos, de esta forma los pasos son concretos. Las fases que recoge son: (1) recepción, (2) autenticación, (3) examen, (4) recolección, (5) reducción, (6) identificación, (7) clasificación, (8) organización, (9) comparación, (10) hipótesis, (11) análisis, (12) atribución, (13) evaluación, (14) interpretación, (15) reconstrucción, (16) comunicación y (17) revisión. A continuación, se detallan cada uno de los puntos:

1. Recepción: esta fase consiste en recibir la prueba y asegurarla. Para ello se hacen dos copias binarias asegurando que la prueba original no sea para nada modificada. El trabajo se realizará con una de las copias mientras que la otra copia deberá ser almacenada para poder disponer, en caso de necesidad, de una copia idéntica a la prueba original.
2. Autenticación: para que el trabajo tenga validez debe asegurarse la autenticidad de las copias con las que se trabajan. Para ello, se usará una función *hash* para la prueba original y también para la copia asegurando que el resultado del *hash* sea el mismo para el original y

la copia. Es prácticamente imposible que teniendo el mismo *hash* el contenido no sea exactamente el mismo y de esta forma podemos demostrar la validez del proceso.

3. Examen: esta es la parte del proceso en la que el investigador consigue extraer los datos de forma legible para los humanos. Para ello, se deberá recuperar todo tipo de datos que estén borrados u ocultos haciendo uso de herramientas forenses.
4. Recolección: en esta fase se reúnen y organizan todos los datos obtenidos en la fase anterior. Durante esta fase, se debe poner de manifiesto cuáles han sido los datos encontrados que habían sido borrados o estaban ocultos.
5. Reducción: es necesario reducir la cantidad de datos recogidos durante el examen, ya que pueden ser muchos y la mayoría no serán de utilidad. Un ejemplo de esto son todos los datos de *logs* conocidos del sistema que no aportarán nada a la investigación. Tras este paso deben quedar solamente los datos que sean atribuibles al usuario.
6. Identificación: esta fase es un paso más para filtrar los datos necesarios. Esta vez se deben localizar aquellos datos que guarden relación con el hecho investigado. Llegados a este punto, el volumen de datos a analizar debe haberse reducido lo suficiente como para poder trabajar con ellos.
7. Clasificación: este punto consiste solamente en agrupar los distintos datos que quedan según el patrón deseado. Por ejemplo, puede ser interesante agrupar los datos por fecha o por tipo de archivos.
8. Organización: en esta fase se trata de intentar agilizar el proceso centrandolo en aquellos tipos de datos que sean claves según el incidente que se trate. Los datos deben ser reestructurados siguiendo este criterio.
9. Comparación: el objetivo de este paso es comparar los datos obtenidos con otros datos obtenidos en otras investigaciones. Realizando esta comparación se pueden llegar a conclusiones basándose en experiencias pasadas.
10. Hipótesis: con los datos obtenidos ya se puede hacer una primera hipótesis. El objetivo de esta hipótesis es el de detectar la causa del incidente y tener un punto de trabajo de cara al posterior análisis.
11. Análisis: esta fase consiste en estudiar los datos relevantes que puedan ayudar a confirmar la hipótesis formulada. Es importante destacar que, legalmente, en esta fase se deben analizar solamente aquellos datos que guarden relación con el caso.
12. Atribución: se deben poder atribuir a alguien los datos recolectados. Para que la prueba tenga validez debe poder ser relacionada inequívocamente con su autor.
13. Evaluación: después de analizar las pruebas, se deben evaluar para determinar si la hipótesis propuesta anteriormente es consistente con los datos hallados.
14. Interpretación: cuando los resultados indican que las pruebas dan validez a la hipótesis, es momento de interpretar de qué modo se deben usar las pruebas para que sean aceptadas legalmente.
15. Reconstrucción: una vez concluida la interpretación de

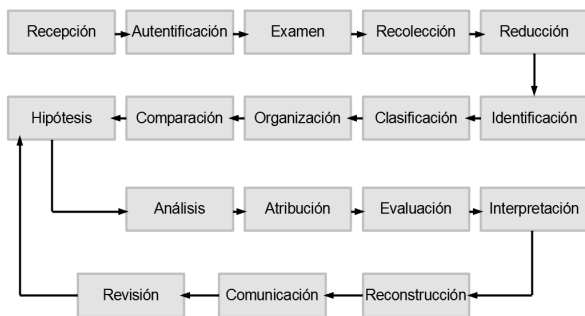


Figura 2. Esquema del modelo 2

los hechos, es momento de reconstruir estos hechos apoyándose en las pruebas de forma que quede un relato probado de la secuencia temporal.

16. Comunicación: en esta fase, la prueba y las conclusiones obtenidas deben ser entregadas a las autoridades competentes. En la mayoría de los casos será entregado a quien dio las pruebas y encargó el trabajo.
17. Revisión: en este último punto se repasará el trabajo hecho en busca de mejoras de cara a una posible presentación y documentación del caso. Llegados a este punto, puede ser que la revisión no sea satisfactoria y se estime necesario volver al punto 10 para volver a formular la hipótesis y seguir con los siguientes pasos teniendo en cuenta esta nueva hipótesis. Una vez concluida la revisión, ya se puede realizar el informe definitivo.

Este comportamiento viene ilustrado en la Fig. 2.

III-C. Modelo 3

Digital Forensics in Social Networks and the Cloud del año 2015 [7]. Este modelo se centra en evaluar el análisis forense para dos grandes retos actuales como son el *cloud computing* y las pruebas que pueden recogerse en las redes sociales. El modelo se compone de cuatro pasos: (1) recolección de la prueba, (2) extracción de la prueba, (3) análisis de la prueba y (4) reparación de la prueba. Estos pasos están ilustrados en la Fig. 3. Seguidamente se explican estos puntos:

1. Recolección de la prueba: la recolección de la prueba debe hacerse en la propia nube, así que puede ser necesario ponerse en contacto con el proveedor del servicio de la nube para poder guardar o recuperar información. Muchas veces será necesario actuar desde el terminal local ya que los datos no serán accesibles desde cualquier punto de internet. En esta fase también es importante estudiar los nodos que están comunicados con las pruebas para poder extraer más información.
2. Extracción de la prueba: en este punto se deben usar herramientas de forense para recuperar datos borrados u ocultos. No se especifica un procedimiento que sea específico para trabajar con la nube. Esto se debe a que todavía queda mucho trabajo que hacer sobre la nube como se pone de manifiesto en [8], [9].
3. Análisis de la prueba: el análisis se basa en dos aspectos que son la limpieza de datos redundantes y el análisis profundo de los datos. Los datos recolectados en este tipo de extracciones contienen mucha información redundante así que, para asegurar la utilidad de las pruebas, se deben analizar solamente los datos importantes. El análisis en profundidad de los datos importantes puede incluir la comparación de las pruebas con la de otros casos que puedan servir de referencia.



Figura 3. Esquema del modelo 3

4. Reparación de la prueba: para asegurar la validez de la prueba es necesario no provocar cambios en el entorno de la prueba. Esto puede demostrarse haciendo uso de un *timestamp* junto con un *hash*. Hacer una copia espejo o usar simulaciones pueden ayudar a reproducir el caso sin alterar el entorno de la prueba.

Este modelo viene ilustrado en la Fig. 3.

III-D. Modelo 4

Research on Live Forensics in Cloud Environment del año 2013 [10]. Este modelo se centra en el estudio forense en vivo del *cloud computing*. El modelo consta de cuatro fases: (1) identificación, (2) recolección, (3) adquisición y (4) preservación. En este modelo se da la opción de escoger entre los puntos (2) y (3) de forma que se puede trabajar usando recolección o adquisición tal y como se muestra en la Fig. 4. A continuación, se detallan las fases del modelo:

1. Identificación: esta fase consiste en identificar cuáles son los datos que serán útiles en la investigación. Además, cuando se está trabajando con la nube es necesario identificar qué tipo de servicio es el que está ofreciendo la nube. Los distintos tipos de nube son SaaS, PaaS e IaaS y en cada uno de ellos se tiene que identificar a su vez otros elementos que pueden ser de utilidad. Con SaaS se pueden identificar, entre otros, todos los ordenadores conectados a la red de clientes. Con PaaS se pueden identificar elementos como los registros disponibles de la aplicación o advertencias del sistema operativo. Con IaaS se pueden identificar elementos como copias de seguridad o imágenes de la máquina virtual.
2. Recolección: en este punto se tratará de conseguir extraer las pruebas desde la nube hasta el laboratorio o un entorno controlado para realizar su análisis.
3. Adquisición: la adquisición consiste en realizar una copia del contenido de la nube para poder trabajar directamente con la copia. Debido a la naturaleza de la nube, normalmente este será el paso que deberá hacerse en lugar de la recolección expuesta anteriormente. La adquisición debe centrarse en los datos digitales y no en los contenedores físicos donde están almacenados.
4. Preservación: esta fase consiste en proteger la integridad de la prueba. Debe evitarse que la prueba sea manipulada o estropeada y esto incluye proteger el entorno de la nube. Además la cadena de custodia debe ser mantenida igualmente y esto trabajando con la nube es más complicado ya que las pruebas pueden estar conectadas a distintos países con distintos marcos jurídicos.

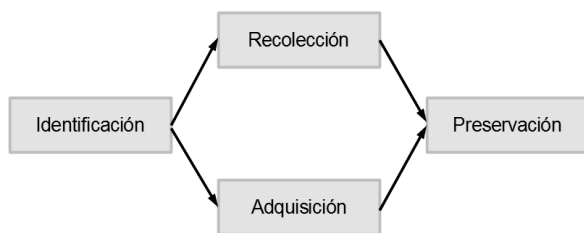


Figura 4. Esquema del modelo 4

IV. COMPARACIÓN DE MODELOS

En esta sección compararemos los modelos introducidos en la sección III. Existen entre ellos muchas similitudes y unas pocas diferencias. Todos ellos comparten un esqueleto común formado por identificación, preservación y análisis de las evidencias recogidas. Aunque algunas de estas fases no aparezcan como tales en algunos de los modelos descritos, son fundamentales para diseñar un modelo de procedimiento adecuado para el análisis forense.

La identificación es tratada como paso previo al trabajo de laboratorio y en los modelos se recoge la necesidad de marcar un camino previo que marque como se desarrollará la investigación. Para ello, es fundamental estudiar en profundidad todos los factores que rodean al caso y que puedan ayudar a enfocar, por ejemplo, qué datos vale la pena analizar. Sin embargo, en el modelo 2 la identificación no se encuentra al principio del proceso sino más adelante. Una vez recuperados todos los datos, se procede con la identificación para quedarse con los datos que guarden relación con el caso. El hecho de no poner la identificación al principio es aceptable ya que lo realmente importante es que se haga antes del análisis y puede que no sea necesaria antes de la extracción de la prueba.

La preservación de las evidencias debe ser una parte importante del proceso forense y los modelos la abarcan de distinta manera. Para algunos, se considera una fase independiente. Este es el caso del modelo 4. En otros de los modelos bajo estudio, se trata de una fase transversal que abarca todos los puntos. Este es el caso del modelo 1, donde la preservación de las evidencias es eje fundamental sin que tenga ningún punto dedicado específicamente a ella. En los modelos se hace hincapié en tener mucho cuidado con las posibles interferencias que puedan producir elementos externos que puedan contaminar la prueba. En el caso de trabajar con la nube, se insiste en proteger el entorno de la nube ya que puede repercutir de forma importante en la prueba original.

El análisis de las evidencias recogidas es el núcleo de todos los modelos, aunque según el modelo este paso se limita solamente al estudio de los datos importantes o puede llegar a englobar todo el proceso de recuperación de datos y como se filtran. Por ejemplo, en el modelo 2 los pasos están muy divididos de forma que no haya ambigüedad sobre ellos, de esta forma cada parte del análisis tiene su propio momento dentro del procedimiento.

Precisamente la importancia que se le da a la especificación de cada paso del análisis es una de las diferencias más notables entre los modelos. Al contrario que el modelo 2, los modelos 3 y 4 son muy generales y no abordan muy profundamente como se realizan las técnicas forenses. Esto

es debido a que estos modelos abarcan también el análisis forense en el ámbito aún abierto del *cloud computing* y algunos pasos se deben volver a definir a partir de trabajos futuros. Pero a parte de las diferencias técnicas de cada paso, las diferencias también están en las especificaciones del proceso en general. Es decir, el análisis forense dentro del modelo se puede quedar en las partes que influyen en el trabajo de laboratorio o ir más allá, mostrando sus pasos previos y posteriores. De esta forma, encontramos modelos que hablan del trabajo en la escena del incidente y que también ponen el foco de atención en la exposición posterior del estudio y su implicación en un juicio.

La fase de documentación es el objetivo final de todo el análisis forense. La documentación es recogida de forma distinta según el modelo. Puede formar parte de uno de los últimos puntos, como es el caso del modelo 1, o bien puede estar indicada en distintos puntos durante todo el proceso definiéndola como fase transversal. En cualquier caso, todos los modelos estudiados acaban con la presentación de un informe final con las conclusiones del estudio forense.

Finalmente, es destacable ver como los modelos no son siempre totalmente lineales. Hemos visto como en el modelo 2 se especifica que los últimos pasos se pueden hacer en bucle hasta encontrar las conclusiones deseadas. Tampoco es exactamente lineal el modelo 4, que da la opción de realizar un punto u otro según el tipo de trabajo. Además, comparando modelos podemos ver que el orden tampoco es algo completamente rígido ya que hay diferentes posibilidades a la hora de ordenar las fases.

V. CONCLUSIONES

Año a año se presentan nuevos modelos de análisis digital forense. Aunque la naturaleza de los delitos varíe poco, las tecnologías y por tanto las evidencias forenses que pueden recogerse sí que lo hacen. Los continuos avances tecnológicos exigen una revisión constante de los modelos para dar cabida a los nuevos retos propuestos.

Los modelos estudiados en este artículo pueden tener poco uso o quedarán obsoletos de forma rápida, pero su estudio nos ayuda a mejorar y profundizar el trabajo del análisis forense digital. Cada nuevo modelo ayuda en la elaboración de los siguientes, sirviendo de referencia de los siguientes modelos durante años. Desarrollando modelos continuamente se tienen referencias recientes del trabajo que se está haciendo de modo que puedan ayudar en otros trabajos.

Actualmente, uno de los retos más importantes del estudio forense digital es que se dedica a trabajar con el *cloud computing* y cómo filtrar las evidencias interesantes dentro de la enorme cantidad de información que se puede recoger y almacenar en los sistemas informáticos tras cada incidente. Los modelos más recientes recogen esa necesidad pero aún les falta mucho recorrido debido al mucho trabajo que queda por hacer en este campo.

Otra de las ventajas de los modelos es que nos ayudan a organizar el trabajo de forma estructurada. Esto es especialmente útil a la hora de redactar un informe del caso ya que una exposición ordenada y bien organizada puede ayudar a defender mejor la postura que se quiera defender.

Utilizar un modelo correcto y sistemático durante un análisis digital forense nos permite proceder de forma adecuada,

adaptarnos perfectamente a las necesidades de cada momento y proveer de un informe final con la rigurosidad y repetitividad necesarias para ser aceptado en un procedimiento legal.

AGRADECIMIENTOS

Este proyecto ha sido financiado parcialmente por el gobierno de España a través del proyecto TAMESIS (TEC2011-22746).

REFERENCIAS

- [1] Mark Reith, Clint Carr, Gregg Gunsch, "And Examination of Digital Forensics Models.", *International Journal of Digital Evidence*, vol. 1, Issue 3, Fall 2002.
- [2] Felix C. Freiling, Bastian Schwittay, "A Common Process Model for Incident Response and Computer Forensics. ", 2007.
- [3] Ankit Agarwal, Megha Gupta, Saurabh Gupta, Subhash Chandra Gupta, "Systematic Digital Forensic Investigation Model. ", 2011.
- [4] Marcus K. Rogers, James Goldman, Rick Mislán, Timothy Wedge, Steve Debrota, "Computer Forensics Field Triage Process Model ", *Journal of Digital Forensics, Security and Law*, Vol. 1, 2006.
- [5] Shahzad Saleem, Oliver & Popov, Ibrahim Bagilli, " Extended abstract digital forensics model with preservation and protection as umbrella principles ", 18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, 2014.
- [6] Michael Donovan Köhn, " Integrated Digital Forensic Process Model ", University of Pretoria, 2012.
- [7] Lei Chen, Lanchuan Xu, Xiaohui Yuan, Narasimha Shashidhar, "Digital Forensics in Social Networks and the Cloud", International Workshop on Sensor, Peer-to-peer and Social Networks, ICNC Workshop, 2015.
- [8] Dominik Birk "Technical challenges of forensic investigations in cloud computing environments." Workshop on cryptography and security in clouds, 2011.
- [9] Draft NISTIR 8006, "NIST Cloud Computing Forensic Science Challenges", 2014.
- [10] Yunting Lei, Yuyin Cui, "Research on Live Forensics in Cloud Environment" 2nd International Symposium on Computer, Communication, Control and Automation, 2013.

Primeros resultados hacia la detección automática de bots en Twitter

Luis Sánchez Fernández, Tello Miñana Rontomé,
Jesús Arias Fisteus, Pablo Basanta Val,
Damaris Fuentes-Lorenzo, Mariluz Congosto
Departamento de Ingeniería Telemática,
Universidad Carlos III de Madrid
Av. Universidad, 30, E-28911 Leganés (Madrid).
webtlab@inv.it.uc3m.es

Norberto Fernández García
Centro Universitario de la Defensa
Escuela Naval Militar
Plaza de España, s/n, 36920 Marín (Pontevedra)
norberto@tud.uvigo.es

Resumen—Las redes sociales en línea gozan de una gran popularidad. A su vez, es bien conocido que parte de la información que se difunde en dichas redes es, en muchos casos, de baja calidad, rumores o simplemente información falsa, difundida frecuentemente por cuentas automáticas o bots (aquellas que son gestionadas por programas de ordenador). En esta comunicación se presentan los primeros resultados de un trabajo que se está realizando para la detección de cuentas bot en Twitter. Entre los rasgos innovadores de este trabajo está el de utilizar indicadores sintácticos para analizar cuentas de Twitter.

Palabras Clave—Redes sociales, Twitter, calidad de fuentes de información, bots

I. INTRODUCCIÓN

Las redes sociales en línea [1] están suponiendo una revolución en la difusión y crecimiento de la información accesible a través de Internet adicional a la que supuso el propio desarrollo de la Web [2], [3]. Al igual que pasa con la Web, la naturaleza abierta de estas redes es una de las causas de su éxito. Sin embargo, también es una de sus debilidades: cualquiera puede difundir información en una red social y no hay ninguna garantía de que dicha información sea verídica [4], [5], [6]. En otras ocasiones, las redes sociales son utilizadas por organizaciones y empresas como medio para intentar influir en la opinión pública [7], [8], [9] o con fines publicitarios [10]. Desafortunadamente, también pueden darse situaciones en que las redes sociales sean una herramienta en la comisión de delitos [11].

En diferentes ámbitos (búsqueda de información [12], protección contra mensajes basura (spam) [13], seguridad [11], análisis de opinión [14], salud pública [15], [16], etc.), es interesante disponer de herramientas que puedan ayudar a evaluar la (buena o mala) calidad de fuentes de datos de redes sociales. Dentro de este marco, diferenciar las cuentas de usuarios humanos de aquellas que tienen detrás a programas de ordenador (llamadas bots) es interesante por varios motivos. Por una parte, en muchos casos los bots son productores de spam [17]. Por otra parte, cuando el objetivo es identificar usuarios humanos expertos en determinado tema es necesario desechar las cuentas bot.

Esta comunicación presenta los resultados de algunos experimentos que se han realizado con datos tomados de un conjunto de cuentas registradas en la red social Twitter, con el objetivo de detectar cuentas bot.

El resto del artículo está organizado de la siguiente forma. La sección II enuncia un conjunto de hipótesis cuya veracidad permitiría diferenciar automáticamente cuentas bot de cuentas de personas con una tasa de error baja. La sección III describe la metodología que se va a utilizar para validar dichas hipótesis. La sección IV describe los tipos de cuentas Twitter que se han analizado, así como las principales fuentes que se han utilizado para su selección. La sección V presenta los resultados que se han obtenido en los experimentos realizados. La sección VI presenta una somera discusión de dichos resultados. La comunicación termina presentando algunos trabajos relacionados y detallando las conclusiones obtenidas en este estudio, así como posibles líneas futuras de continuación del mismo.

II. HIPÓTESIS DE PARTIDA

La cuestión de distinguir el comportamiento de un ser humano del de una máquina es muy antigua. Un artículo clásico al respecto es el famoso “Computing machinery and Intelligence”, de Alan Turing [18] (en realidad este trabajo plantea la cuestión inversa: si es posible construir una máquina cuyo comportamiento sea indistinguible del de los seres humanos).

Dado que nuestro primer objetivo, en el que nos centramos en esta comunicación, es el de ser capaces de distinguir automáticamente las cuentas Twitter de personas de las cuentas bot con una tasa de acierto elevada, el primer paso es construir de forma razonada algunas hipótesis cuya validez pudiera permitir diferenciar las cuentas bot de las que no lo son.

En primer lugar vamos a intentar identificar diferencias en el patrón temporal de tuiteo de las cuentas bot y de las cuentas de personas. Es bien conocido que el comportamiento temporal de los seres humanos presenta una importante variabilidad [19]. Es natural pensar que el comportamiento temporal de los programas de ordenador sea más regular (salvo que esté específicamente diseñado para no serlo). Esta idea ya ha sido explorada recientemente en trabajos como [20] y será la base de la primera de nuestras hipótesis.

En segundo lugar, nos planteamos explotar la dificultad de sintetizar lenguaje natural de forma automática [21]. Aparte de la complejidad intrínseca del lenguaje natural, es bien conocido que los seres humanos rompemos continuamente las reglas gramaticales del idioma en el que nos expresamos [21].

En resumen, al igual que en el caso anterior, esperamos encontrar una mayor variabilidad en el lenguaje expresado por los humanos que en el sintetizado automáticamente por un programa de ordenador.

Conviene en este caso matizar que existen experimentos previos que muestran que los programas de ordenador pueden de forma bastante efectiva producir lenguaje natural cuando mantienen un diálogo con un ser humano. En este contexto el programa de ordenador explota las sentencias generadas por el humano para producir las suyas propias. Uno de los primeros experimentos que se realizaron en esta línea fue el programa ELIZA [22]. De hecho, los bots conversacionales pueden ser considerados como una categoría de bots de Twitter en sí misma.

La tercera hipótesis que vamos a plantear es específica del ámbito de las redes sociales en línea. Como ya se ha dicho, muchos de los bots se utilizan para difundir spam. En otros casos, se utilizan bots para difundir artículos de blogs y otras fuentes relacionados con diferentes temas. En todos estos casos es de esperar que los tuits de los bots contengan enlaces en un porcentaje elevado de casos, mientras que los humanos utilizaran enlaces con menos frecuencia.

En resumen de la discusión anterior, pretendemos evaluar las siguientes tres hipótesis:

H1 Las cuentas bot tuitean de forma regular, mientras que los humanos pueden estar largos intervalos de tiempo sin tuitear.

H2 Los tuits de las cuentas bot utilizan menos variedad en el lenguaje que los usuarios humanos.

H3 Los tuits de las cuentas bot contienen enlaces con una frecuencia mucho mayor que los tuits de las cuentas de personas.

En la sección siguiente se describe la metodología utilizada para intentar evaluar la validez de estas hipótesis.

III. METODOLOGÍA UTILIZADA

Para evaluar las tres hipótesis de partida presentadas en la sección anterior, se han definido los siguientes cuatro indicadores para una cuenta de Twitter:

- 1) Desviación típica de la diferencia entre segundos entre el sello temporal de un tuit y su anterior. La razón por la que se ha escogido este indicador es que si una cuenta tuitea de forma regular tendrá una desviación típica más baja que las cuentas en las que los intervalos de tuiteo sean irregulares. Relacionado con hipótesis **H1**.
- 2) Intervalo máximo entre un tuit y el anterior. Se espera que los bots tengan valores menores de este indicador porque tuitean de forma más regular. Relacionado con hipótesis **H1**.
- 3) El siguiente indicador está relacionado con la hipótesis **H2**. En relación con dicha hipótesis podrían contemplarse múltiples indicadores. En esta primera aproximación nos hemos conformado con uno solo. Partiremos de que las cuentas bot utilizan sentencias con una estructura sintáctica similar en sus tuits. Para intentar medir esto vamos a medir la variabilidad (en términos de desviación típica) de la profundidad del árbol sintáctico de cada sentencia escrita en una cuenta de Twitter. Para obtener dichos árboles sintácticos hemos utilizado el parser de la Universidad de Stanford [23].

- 4) Porcentaje de tuits que contienen enlaces. Derivado directamente de la hipótesis **H3**.

Para evaluar la validez de las hipótesis de partida, se han medido estos indicadores para un conjunto de cuentas en Twitter suficientemente grande como para que los resultados sean estadísticamente significativos y de las que se conoce a priori si son bots o no. A continuación los valores de los indicadores obtenidos para cada tipo de cuenta se han presentado en forma de box-plot, tal y como se describe en la sección de resultados.

Con el objetivo de que los datos recogidos en este trabajo puedan ser utilizados en otros experimentos posteriores, las cuentas no se han clasificado exclusivamente entre cuentas bot o no, tal y como se detalla en la sección siguiente.

IV. RECOLECCIÓN DE DATOS

Para realizar los experimentos se seleccionaron cuentas de Twitter de los siguientes tipos:

- Cuentas de medios de comunicación, todas ellas verificadas oficialmente por Twitter y obtenidas de [24].
- Cuentas de investigadores, profesores de universidad y periodistas, obtenidas de los seguidores de cuentas de universidades prestigiosas del mundo como Stanford y Cambridge.
- Cuentas de personajes famosos (políticos, actores, deportistas, etc.) y verificadas.
- Cuentas “fake”: cuentas que parodian a los personajes de la categoría anterior, la mayoría de ellas obtenidas de [25], [26], [27].
- Cuentas bot, obtenidas de [28] y otras por medio de búsqueda manual.
- Cuentas aleatorias. La idea de esta categoría es disponer de cuentas de usuarios normales que no pertenezcan a ninguna de las categorías anteriores. Para obtener las cuentas aleatorias nos conectamos al Stream API de Twitter [29] y escogimos las primeras 40 cuentas en inglés y que hayan tuiteado al menos 10 veces desde que se creó la cuenta para tener algunos datos con los que poder realizar el análisis.

Para cada tipo de cuenta de Twitter se han analizado 40 cuentas. Esto debe ser suficiente, por ejemplo, para entrenar un clasificador con la técnica clásica del análisis discriminante de Fisher [30], [31] utilizando muestras de 20 dimensiones [32].

Como tenemos 6 tipos de cuentas y 40 cuentas de cada tipo, en total se han analizado 240 cuentas.

V. RESULTADOS OBTENIDOS

Cada uno de los cuatro indicadores descritos en la sección III se ha calculado para cada una de las 240 cuentas consideradas. Los resultados obtenidos se presentan por medio de un diagrama de cajas o box-plot [31], [33].

La figura 1¹ representa un diagrama de caja. Un diagrama de caja es un gráfico, basado en cuartiles, mediante el cual se visualiza un conjunto de datos. Está compuesto por un

¹“Boxplot” de Jumanbar - Trabajo propio. Disponible bajo la licencia CC BY-SA 3.0 vía Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Boxplot.svg#/media/File:Boxplot.svg>

rectángulo, la "caja", y dos brazos, los "bigotes". La información que suministra el diagrama de caja, como se puede ver en la figura 1 es la siguiente:

- Primer cuartil (Q1)
- Segundo cuartil o mediana. Es la raya dentro de la caja.
- Tercer cuartil (Q3)
- Los bigotes, las líneas que se extienden desde la caja, se extienden hasta los valores máximo y mínimo de la serie o hasta 1.5 veces el Rango Inter Cuartílico (Q3 - Q1).
- Cuando los datos se extienden más allá de esto, significa que hay valores atípicos en la serie, que se marcan con pequeños círculos.

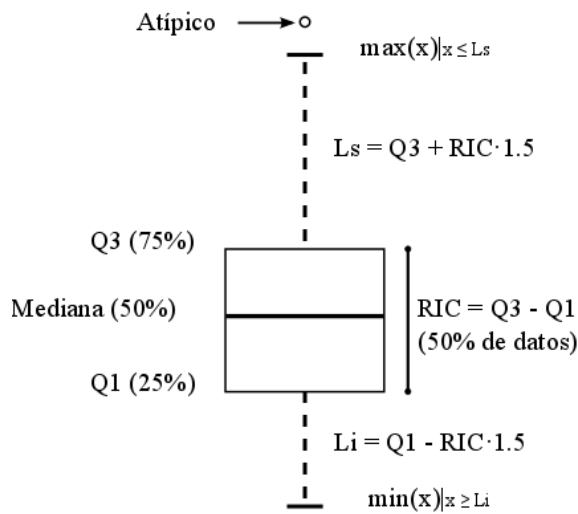


Fig. 1. Diagrama de caja o box-plot

Los resultados obtenidos para cada uno de los indicadores se presentan en las figuras 2, 3, 4 y 5. En cada una de estas figuras, el significado de las etiquetas en el eje de abscisas es el siguiente:

- X1: cuentas de medios de comunicación.
- X2: cuentas bot.
- X3: cuentas fake.
- X4: cuentas de famosos.
- X5: cuentas de investigadores, profesores y periodistas.
- X6: cuentas aleatorias.

Se debe tener en cuenta además que los resultados de las figuras 2 y 3 están expresados en escala logarítmica (logaritmo neperiano).

VI. DISCUSIÓN

Analizando los resultados obtenidos con los indicadores 1 y 2 se concluye que la hipótesis **H1** no es válida o al menos que dichos indicadores no miden adecuadamente la diferencia en el comportamiento temporal de los patrones de tuitos de los usuarios humanos frente a los bots. Sí se observa en estos dos indicadores y especialmente en el indicador 1 una diferencia significativa entre el patrón de tuitos de los medios de comunicación (estas cuentas pueden en muchos casos ser gestionadas total o parcialmente de manera automática) y el de las cuentas de usuarios humanos.

El indicador 3 en cambio confirma la validez de la hipótesis **H2** y de hecho muestra un comportamiento muy diferenciado

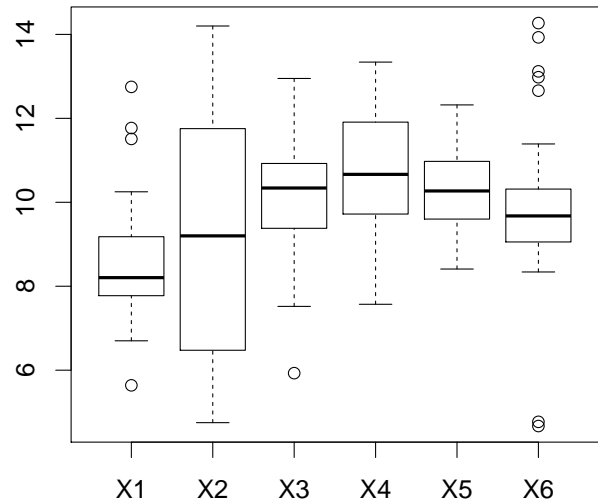


Fig. 2. Resultados para el indicador 1: desviación típica de la diferencia entre segundos entre dos tuits

entre las cuentas bot y las de los demás tipos. Llama la atención particularmente el buen comportamiento de este indicador para diferenciar cuentas bot de las de medios de comunicación, frente a lo ocurrido con los otros tres indicadores. Una posible explicación de este buen comportamiento podría ser la siguiente. Las cuentas Twitter de medios de comunicación habitualmente tuitean los titulares de las noticias que se van generando en el propio medio de comunicación. El texto de estos titulares es generado por periodistas y lógicamente presenta una variabilidad importante. Este indicador funciona mejor que los otros 3 considerados porque aunque las cuentas de medios de comunicación presentan similitudes con las cuentas bot en cuanto a patrón temporal y enlaces son muy diferentes desde el punto de vista de los aspectos lingüísticos de los tuits. En resumen, el indicador 3 se podría considerar adecuado para formar parte de un conjunto de indicadores para la detección de bots.

Finalmente, el indicador 4 también confirmaría la hipótesis **H3** (aunque de forma menos significativa que en el caso anterior). Excluyendo a las cuentas de medios de comunicación (que como ya se ha dicho muestran ciertas afinidades con las cuentas bot), se observa que más de la mitad de las cuentas bot tienen un porcentaje de tuits con enlaces próximo al 100% y notablemente más alto que la inmensa mayoría de las cuentas de los demás tipos. Por lo tanto, este indicador, si bien por sí solo no parece suficiente para distinguir cuentas bot de las que no lo son, parece adecuado para formar parte de un conjunto de indicadores que se utilizasen para detectar cuentas bot.

VII. TRABAJOS RELACIONADOS

Existen varios trabajos relacionados con el estudio de la fiabilidad de la información difundida en redes sociales.

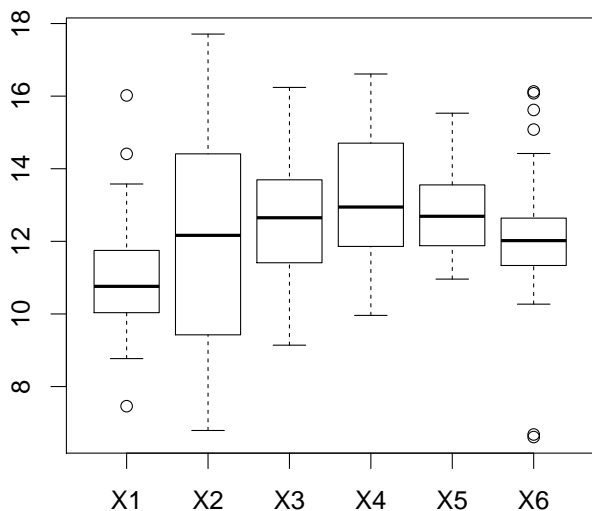


Fig. 3. Resultados para el indicador 2: Intervalo de tiempo máximo entre un tuit y el anterior

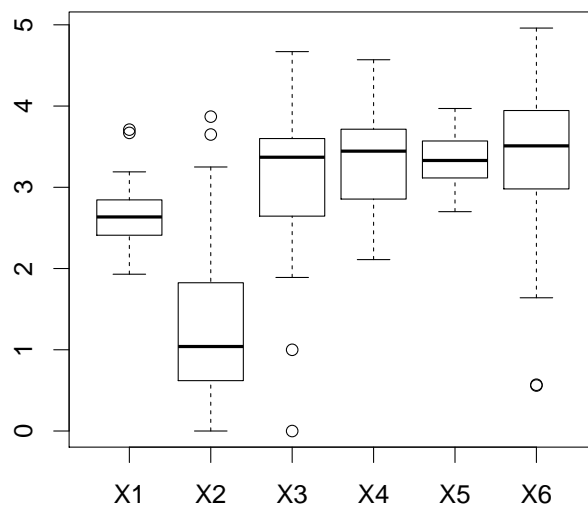


Fig. 4. Resultados para el indicador 3: Desviación típica de la profundidad del árbol sintáctico

En [34] se analiza la fiabilidad de los tuits difundidos durante el desastre de Fukushima. Este estudio se centra en los denominados tuits sintetizados, en los que un usuario sintetiza información proveniente de cierta fuente de información (en contraposición a un retuiteo puro) y que utilizaban el hashtag #fukushima. Se encontró que el 70% de los tuits provenían de fuentes de información fiables, con un porcentaje más bajo cuando los tuits provenían de cuentas con un perfil anónimo.

Otro trabajo en la misma línea es el de [4]. En este trabajo se analizó la difusión de tuits en el terremoto de Chile de 2010. En este estudio se concluyó que los rumores en Twitter tienden a ser más cuestionados que las noticias verídicas, lo que puede ser un instrumento en si mismo para su detección.

La detección de spammers puede considerarse como un apartado dentro del análisis de la fiabilidad de fuentes de información que ha sido estudiado en numerosos trabajos [13], [35], [36], [37].

En cuanto al caso concreto de la detección de bots en Twitter, existen algunos trabajos relacionados con lo aquí expuesto. En [38] se utilizan técnicas de aprendizaje máquina para distinguir spam bots de otros bots. El trabajo presentado en [20] se centra en identificar bots en base a patrones temporales. Principalmente, se han considerado dos aspectos de dichos patrones: 1) la distribución de los intervalos temporales entre 2 tuits; y 2) la frecuencia de tuiteo dependiendo de la hora del día (se debe tener en cuenta que este segundo aspecto solamente es útil cuando se dispone de información acerca de la zona horaria desde la que se tuitea). En [39] se propone una herramienta para la detección de bots. Entre los indicadores considerados en este trabajo se utilizan parámetros temporales. Se concluye que los bots tienden a tuitear en ráfagas, o con patrones temporales con una mayor regularidad

que los humanos. Curiosamente en ese estudio se afirma que los bots no siempre tuitean con más frecuencia que los humanos. Entre otros indicadores también utilizados en este estudio se encuentran los siguientes: clasificación del texto de los tuits como spam, relación entre followers y followees y porcentaje de tuits que contienen enlaces. Finalmente, otra trabajo interesante es el de Ferrara et al. [40] en el que se estudian y clasifican varios enfoques para la detección de bots.

VIII. CONCLUSIONES Y TRABAJO FUTURO

En el estudio que presentamos en este trabajo se han evaluado 4 indicadores con el objetivo de detectar bots de Twitter. Estos 4 indicadores se han diseñado en base a 3 hipótesis sobre las diferencias entre las cuentas bot y las cuentas de usuarios humanos. Los indicadores 1 y 2, relativos al patrón de comportamiento temporal no se han mostrado efectivos por lo que la hipótesis **H1** no ha podido ser validada. En cambio, los resultados obtenidos para los indicadores 3 y 4 son prometedores y parecen validar las hipótesis **H2** y **H3**, relativas a la variabilidad del lenguaje empleado en las cuentas de Twitter y al porcentaje de tuits que contienen enlaces.

Como se ha explicado en la introducción, el estudio que se presenta en este trabajo es un primer paso que debe ser continuado en trabajos posteriores. A continuación se esbozan varias líneas de trabajo que planeamos desarrollar en el futuro. En primer lugar, es necesario ampliar el conjunto de indicadores que se utilizan. Dentro de este apartado se debe repensar la hipótesis **H1** y el uso de indicadores temporales que, como se ha indicado anteriormente no han dado los resultados esperados. Como se ha indicado en la sección de trabajos relacionados, artículos como [39] sugieren un patrón de comportamiento temporal de los bots en ráfagas.

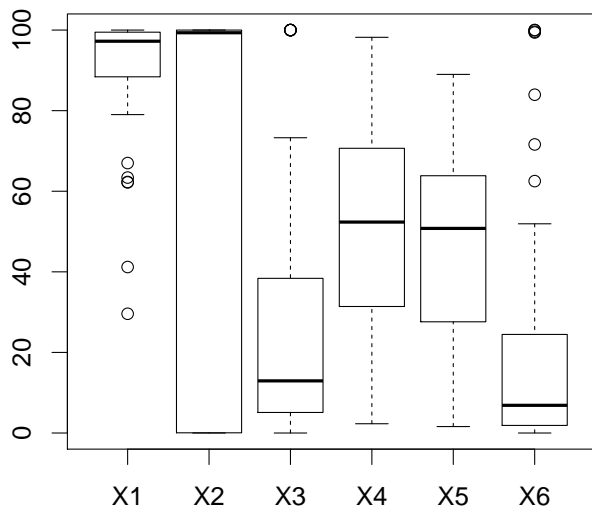


Fig. 5. Resultados para el indicador 4: Porcentaje de tuits que contienen enlaces

Se deben por lo tanto diseñar indicadores que detecten este tipo de comportamiento temporal. Por otra parte, el buen comportamiento observado por el indicador 3 nos anima a explorar el uso de otros indicadores lingüísticos, tanto relativos a la estructura sintáctica de las sentencias contenidas en una cuenta de Twitter como otros relativos al contenido de los tuits.

Una vez que se disponga de un conjunto de indicadores adecuado, será necesario entrenar un clasificador utilizando técnicas de aprendizaje máquina. Como parte de dicho proceso será necesario seleccionar los indicadores que se deben utilizar, para lo cual existen varios modelos estadísticos que se podrían considerar [41], [42]. Finalmente es necesario determinar cual de las múltiples técnicas de clasificación se va a utilizar (support vector machines, análisis discriminante de Fisher, modelo logit, Naive Bayes, etc.).

Finalmente, será necesario realizar una evaluación del clasificador así obtenido utilizando datos de un volumen de cuentas elevado (cientos de miles) para poder garantizar que los resultados obtenidos sean fiables.

La explotación en un entorno realista de una herramienta como la que se pretende desarrollar involucraría el procesado de grandes volúmenes de datos (Big Data) para aplicar estas técnicas a un gran número de cuentas de Twitter. Para tratar con esto sería interesante utilizar modelos predecibles de procesado distribuido de flujos, como el propuesto en [43].

AGRADECIMIENTOS

Este trabajo está parcialmente financiado por el Ministerio de Economía y Competitividad a través del proyecto “HERMES-SMARTDRIVER” (TIN2013-46801-C4-2-R) y

por la Comunidad de Madrid a través del proyecto “eMadrid” (S2013/ICE-2715).

REFERENCIAS

- [1] N. B. Ellison *et al.*, “Social network sites: Definition, history, and scholarship,” *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.
- [2] T. Berners-Lee, M. Fischetti, and M. L. Foreword By-Dertouzos, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper Information, 2000.
- [3] B. A. Huberman and L. A. Adamic, “Internet: growth dynamics of the world-wide web,” *Nature*, vol. 401, no. 6749, pp. 131–131, 1999.
- [4] M. Mendoza, B. Poblete, and C. Castillo, “Twitter under crisis: Can we trust what we rt?” in *first workshop on social media analytics*. ACM, 2010, pp. 71–79.
- [5] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi, “Collective attention in the age of (mis)information,” *Computers in Human Behavior*, no. 0, pp. –, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563215000382>
- [6] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, “Science vs conspiracy: collective narratives in the age of (mis) information,” *PLoS ONE*, vol. 10, no. 2, 2015.
- [7] M. L. Congosto, “Elecciones europeas 2014: Viralidad de los mensajes en twitter,” *Redes: revista hispana para el análisis de redes sociales*, vol. 26, no. 1, pp. 23–52, 2015.
- [8] M. L. Congosto, “Virialidad de los mensajes en twitter en las campañas electorales,” in *III Congreso Internacional en Comunicación Política y Estrategias de Campaña*, 2014.
- [9] C. B. Williams and J. Girish, “Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008,” *New Media & Society*, p. 1461444812457332, 2012.
- [10] J. A. Chevalier and D. Mayzlin, “The effect of word of mouth on sales: Online book reviews,” *Journal of marketing research*, vol. 43, no. 3, pp. 345–354, 2006.
- [11] C. C. Yang and T. D. Ng, “Terrorism and crime related weblog social network: Link, content analysis and information visualization,” in *Intelligence and Security Informatics, 2007 IEEE*. IEEE, 2007, pp. 55–58.
- [12] S. P. Borgatti and R. Cross, “A relational view of information seeking and learning in social networks,” *Management science*, vol. 49, no. 4, pp. 432–445, 2003.
- [13] D. Gayo-Avello, “Nepotistic relationships in twitter and their impact on rank prestige algorithms,” *Information Processing & Management*, vol. 49, no. 6, pp. 1250–1280, 2013.
- [14] F. Li and T. C. Du, “Who is talking? an ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs,” *Decision Support Systems*, vol. 51, no. 1, pp. 190–197, 2011.
- [15] S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, “A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication,” *Journal of medical Internet research*, vol. 15, no. 4, 2013.
- [16] S. A. Adams, “Revisiting the online health information reliability debate in the wake of “web 2.0”: an inter-disciplinary literature and website review,” *International journal of medical informatics*, vol. 79, no. 6, pp. 391–400, 2010.
- [17] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, “Who is tweeting on twitter: human, bot, or cyborg?” in *Proceedings of the 26th annual computer security applications conference*. ACM, 2010, pp. 21–30.
- [18] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [19] A.-L. Barabasi, “The origin of bursts and heavy tails in human dynamics,” *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [20] G. Tavares and A. Faisal, “Scaling-laws of human broadcast communication enable distinction between human, corporate and robot twitter users,” *PloS one*, vol. 8, no. 7, p. e65774, 2013.
- [21] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach (3rd edition)*. Prentice Hall, 2009.
- [22] J. Weizenbaum, “Eliza-a computer program for the study of natural language communication between man and machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
- [23] D. Klein and C. D. Manning, “Accurate unlexicalized parsing,” in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003, pp. 423–430.
- [24] N. Bremmen, “The 100 most influential news media twitter accounts,” <http://memeburn.com/2010/09/the-100-most-influential-news-media-twitter-accounts/> (visitado el 7 de mayo de 2015), 2010.

- [25] C. McCann, "The list: Ten fake twitter accounts," <http://www.ft.com/cms/s/2/c843804c-3b21-11e2-b3f0-00144feabdc0.html#axzz2jz08haZW> (visitado el 7 de mayo de 2015).
- [26] I. Paul, "15 fake and funny twitter accounts," http://www.pcworld.com/article/159492/fake_funny_twitter.html (visitado el 7 de mayo de 2015), 2009.
- [27] E. Levine, "The 15 best fake twitter accounts of 2012," <http://heavy.com/comedy/2012/12/the-15-best-fake-twitter-accounts-of-2012/> (visitado el 7 de mayo de 2015), 2012.
- [28] K. Knibbs, "The 8 best twitter bots you aren't following," <http://www.digitaltrends.com/social-media/the-10-best-twitter-bots-you-arent-following/> (visitado el 7 de mayo de 2015), 2013.
- [29] "The streaming apis," <https://dev.twitter.com/streaming/overview> (visitado el 7 de mayo de 2015), 2015.
- [30] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate analysis*. Academic press, 1979.
- [31] D. Peña, *Análisis de datos multivariantes*. Mc Graw Hill Interamericana de España, S.A.U., 2002.
- [32] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: Recommendations for practitioners," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 3, pp. 252–264, 1991.
- [33] M. Lavine, "Introduction to statistical thought," <http://people.math.umass.edu/~lavine/Book/book.html> (visitado el 7 de mayo de 2015), 2005.
- [34] R. Thomson, N. Ito, H. Suda, F. Lin, Y. Liu, R. Hayasaka, R. Isochi, and Z. Wang, "Trusting tweets: The fukushima disaster and information source credibility on twitter," in *9th International ISCRAM Conference*, 2012, pp. 1–10.
- [35] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *26th Annual Computer Security Applications Conference*. ACM, 2010, pp. 1–9.
- [36] A. H. Wang, "Don't follow me: Spam detection in twitter," in *Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on*. IEEE, 2010, pp. 1–10.
- [37] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," in *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*, vol. 6, 2010, p. 12.
- [38] A. H. Wang, "Detecting spam bots in online social networking sites: a machine learning approach," in *24th Annual IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy*. Springer, 2010, pp. 335–342.
- [39] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *Dependable and Secure Computing, IEEE Transactions on*, vol. 9, no. 6, pp. 811–824, 2012.
- [40] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *arXiv preprint arXiv:1407.5225*, 2014.
- [41] G. Schwarz *et al.*, "Estimating the dimension of a model," *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [42] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*. Akademinai Kiado, 1973, pp. 267–281.
- [43] P. Basanta, N. Fernandez-Garcia, A. Wellings, and N. Audsley, "Improving the predictability of distributed stream processors," *Future Generation Computing Systems*, 2015.

Análisis de Datos Funcionales para Gestión de Red: Técnicas, Retos y Oportunidades

David Muelas*, Jorge E. López de Vergara*, José R. Berrendero[†], Javier Aracil*

*Departamento de Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior

[†]Departamento de Matemáticas, Facultad de Ciencias

Universidad Autónoma de Madrid

Email: {dav.muelas, jorge.lopez_vergara, joser.berrendero, javier.aracil}@uam.es

Resumen—Las tareas de gestión de red se caracterizan actualmente por la diversidad de las variables monitorizadas y de las situaciones que deben ser consideradas. Las nuevas arquitecturas de red inducen además cambios cada vez más rápidos, produciendo un aumento de la volatilidad del comportamiento de las infraestructuras. Teniendo en cuenta este contexto, es necesario que las herramientas y métodos aplicados a las tareas de gestión se adecuen y evolucionen para superar las nuevas necesidades y restricciones. Además, las capacidades actuales en términos de procesamiento de fuentes heterogéneas y masivas de datos abren la puerta a la introducción de nuevas aproximaciones. En este trabajo proponemos la aplicación de las técnicas de Análisis de Datos Funcionales (*Functional Data Analysis*, FDA) como alternativa que supere las limitaciones de los métodos existentes y aproveche las nuevas capacidades de los sistemas de análisis de datos. Ilustraremos cómo estas técnicas proporcionan medios para encarar las distintas etapas del análisis de datos de red, incluyendo herramientas que pueden ser aplicadas a las distintas áreas funcionales de la gestión de red.

Palabras Clave—gestión de red, análisis de datos, modelado de tráfico, líneas base, detección de anomalías, FDA, PCA funcional, profundidad

I. INTRODUCCIÓN

Las técnicas de análisis de datos aplicadas en las labores de gestión de red están inmersas en un continuo cambio como consecuencia de la evolución de las necesidades y condiciones de operación de las arquitecturas de red. Este factor, unido a los cambios cada vez más rápidos de la dinámica de las redes, hacen que si las soluciones analíticas aplicadas carecen de una flexibilidad y adaptabilidad suficiente puedan resultar inútiles en las nuevas infraestructuras de telecomunicación. Por ejemplo, las aproximaciones que requieren una cierta estructura de las observaciones (*e.g.* gaussianidad de los datos) pueden no ser utilizables en despliegues que no presentan las características requeridas (tal es el caso de las situaciones descritas en [1], [2]). Esto nos lleva a que, además de tener un coste computacional razonable, las nuevas herramientas de análisis para la gestión de red deben ser flexibles y adaptables.

Si a esto le sumamos los avances en las arquitecturas de almacenamiento y procesamiento de datos propias de la era del *Big Data*, resulta necesario que se sigan abriendo nuevas líneas de investigación en la gestión de red para optimizar los procesos de toma de decisiones en esta área. Con las características actuales de estos sistemas, no obstante, conviene optimizar la cantidad de datos almacenados. La latencia puede ocasionar la pérdida de datos vitales en entornos de altas prestaciones si no se utilizan técnicas de preprocesado que mejoren la escritura de grandes volúmenes de datos en tiempo real. Así, la selección y compresión de registros informativos

que representen fielmente el estado de la red es un problema capital, especialmente de cara a permitir el análisis forense a largo plazo. Además, es necesario explorar soluciones que, siendo impracticables en otro momento debido a las limitaciones tecnológicas, puedan permitir en la actualidad una mejor comprensión del comportamiento de las redes de comunicaciones.

Por otro lado, la preocupación por el anonimato de los datos de red no deja de aumentar. El cifrado de las comunicaciones y otros aspectos legales relativos a estos datos limita algunas soluciones existentes en el estado de la técnica. Como situación que ilustra la importancia de este hecho, podemos ver que sistemas de seguridad o de control de tráfico que requieran de la Inspección Profunda de Paquetes (*Deep Packet Inspection*, DPI) pueden no ser aplicables en situaciones en las que sea necesario garantizar un cierto nivel de seguridad y anonimato.

Por todo lo expuesto, proponemos la aplicación de las técnicas de Análisis de Datos Funcionales (*Functional Data Analysis*, en adelante FDA) a las tareas involucradas en las labores de gestión y análisis de redes de comunicaciones. FDA se basa en la consideración de variables aleatorias con valores en espacios funcionales, lo que supone el estudio y aplicación de herramientas estadísticas en espacios de dimensión infinita.

En el ámbito de las redes de comunicaciones, existe una gran cantidad de parámetros de rendimiento y funcionamiento que pueden ser considerados como datos funcionales – *e.g.* series temporales o funciones de distribución. La posibilidad (al menos teórica) de realizar observaciones en cualquier instante de tiempo y su característica periodicidad permiten el tratamiento funcional de estos datos [3]. Los buenos resultados que se han obtenido al aplicar FDA a otros problemas con características similares a los de las redes de comunicaciones (*e.g.* predicción de tiempo atmosférico, crecimiento humano y ciertos estudios económicos¹) motivan la exploración de la aplicabilidad de FDA en esta área.

Teniendo en cuenta las cuestiones comentadas previamente y las limitaciones de las soluciones existentes, queremos ilustrar cómo se pueden analizar los datos de red utilizando un enfoque funcional. Con este objetivo en mente, analizaremos técnicas de preprocesado y representación de datos, y distintos métodos que permiten extraer conclusiones de las observaciones realizadas, para proporcionar una visión global de las posibilidades que ofrece esta rama de la estadística actualmente en desarrollo.

¹<http://www.psych.mcgill.ca/misc/fda/examples.html>

Para ilustrar cómo encarar el procesado de datos de red utilizando una aproximación funcional, vamos a restringirnos en este trabajo al procesamiento de series temporales que representan el *throughput* de los datos que atraviesan un punto de presencia de la red, aunque estas técnicas son igualmente aplicables al análisis de cualquier variable representada como serie temporal – e.g. retardos, tiempos de respuesta, etc.

El resto del artículo se estructura de la siguiente forma. En primer lugar, vamos a repasar trabajos previos que presenten soluciones relacionadas con nuestra propuesta, indicando las ventajas que emergen del empleo de FDA. Posteriormente, relacionaremos algunas técnicas de FDA con distintas etapas del procesado de datos en el análisis de red. Proporcionaremos un caso práctico de estudio, mostrando las ganancias del empleo de FDA como conjunto de herramientas de análisis. Finalmente, extraeremos las conclusiones de nuestro estudio e indicaremos algunas líneas de trabajo que estamos empezando a explorar.

II. ESTADO DEL ARTE

Procedemos ahora a comentar algunos resultados existentes que están relacionados con la propuesta que hacemos en este trabajo. Nuestra revisión se plantea en torno a herramientas, métodos y soluciones que bien pueden ser enriquecidos o mejorados con las alternativas basadas en un análisis funcional de los datos de red; bien muestran limitaciones que pueden ser resueltas con el enfoque de FDA.

En primer lugar, consideramos herramientas generales de monitorización de red que pueden ser complementadas con nuestros métodos. Estas herramientas proporcionan asimismo fuentes de datos que pueden tratarse como observaciones funcionales. Las soluciones que comentamos muestran características que ilustran los principios a los que tienden las nuevas aproximaciones para el estudio de redes de comunicaciones. No obstante, en [4], [5] se pueden encontrar revisiones más completas de aplicaciones para la monitorización y obtención de medidas de red.

En [6], los autores presentan *Scap*, un sistema orientado a un análisis basado en flujos de datos y adaptado a redes con alta tasa de tráfico. Este sistema, de acuerdo a la descripción de sus autores, admite el empleo de las técnicas estadísticas desarrolladas en el ámbito de FDA para aumentar sus capacidades y funcionalidad – por ejemplo, incorporando la selección en tiempo real del tráfico que debe ser persistido a partir de la comparación con respecto a líneas base funcionales. *Scap* muestra la importancia creciente de las estadísticas agregadas (e.g. valores proporcionados por SNMP, registros *NetFlow*) en el análisis de redes multi-Gb. La vigencia y empleo de este tipo de estadísticas se debe a que su consideración conlleva una minimización de la carga computacional del análisis de datos, que es un aspecto crítico de estos sistemas.

Otro ejemplo de las herramientas de monitorización propuestas en los últimos tiempos es *BlockMon* [7]. Este sistema, modular y distribuido, ilustra el interés en el desarrollo de arquitecturas de monitorización flexibles para permitir su adaptación a entornos de red heterogéneos con fines de aplicación diferenciados. *BlockMon* puede ser complementado así con módulos que realicen un preprocesado basado en una aproximación funcional para proveer resultados enriquecidos

para el posterior análisis y visualización. Como se verá, las características del análisis de datos funcionales potencia la adaptabilidad del sistema al eliminar ciertas hipótesis que otros métodos imponen sobre las observaciones.

Para terminar la revisión de herramientas de monitorización, mencionamos la propuesta recogida en [8]. En este trabajo, los autores presentan una plataforma de análisis de tráfico de red, *M³OMON*, que proporciona datos con distinta resolución – en particular, series temporales de datos agregados, registros de flujos, y paquetes de red. El trabajo de estos autores ilustra la necesidad de analizar simultáneamente distintas fuentes de datos con distintos niveles de agregación, para detectar correctamente eventos en redes. Sus resultados muestran la capacidad de su sistema para manejar estas situaciones en entornos altamente demandantes, proporcionando un ecosistema de aplicaciones en las que las técnicas basadas en FDA podrían proporcionar nuevas salidas de datos para un análisis posterior en profundidad.

Siguiendo con las tareas involucradas en el flujo de datos en los procesos de análisis, vamos a comentar algunas técnicas de preprocesado de datos de red relacionadas con FDA. Para extender las referencias que vamos a considerar, en [9] se puede encontrar una revisión de aproximaciones específicas existentes en el dominio de la detección de intrusiones.

En [10] los autores proponen el empleo del Análisis de Componentes Principales (PCA) sobre registros de *throughput* para descomponerlos en términos de *eigenflows*. Esta solución comparte algunas similitudes con PCA funcional, que será introducido en la Sección III. No obstante, esta solución no utiliza una representación funcional previa que proporcione una interpretación semántica de las componentes principales. En cualquier caso, las ideas centrales de este trabajo indican las ventajas derivadas del tratamiento funcional de ciertos datos de red.

Otras técnicas de preprocesado que pueden ser vistas como precursoras de FDA son las expuestas en [11], [12]. En estos trabajos, se utilizan técnicas de análisis multi-resolución basado en ondículas para comprimir observaciones de red, con una evaluación de las propiedades estadísticas de los resultados. Formalmente, el análisis multi-resolución proporciona una representación funcional de datos restringida a ciertas bases particulares. Como se verá, la obtención de una representación con respecto a una base funcional es un primer paso habitual al utilizar técnicas de FDA, de forma que los resultados de estos trabajos están incluidos en los casos abordados por nuestra propuesta.

Llegados a este punto, vamos a comentar algunas referencias a trabajos que encaran el modelado de red utilizando soluciones relacionadas con las alternativas funcionales que presentamos.

Los autores de [13] proponen el empleo de una nueva métrica cuyo cálculo conlleva un coste computacional reducido y que condensa información significativa en entornos de monitorización de Centros de Datos. Esta aproximación pone de manifiesto muchos de los principios que guían nuestro estudio, aunque se restringe a un contexto de aplicación muy específico. Una de las ventajas del análisis funcional de datos es que, dadas las reducidas hipótesis que requieren muchos de los métodos, es aplicable a numerosos contextos siempre

y cuando la muestra se ajuste al esquema de observación de funciones. Por ejemplo, este tipo de técnicas se puede aplicar directamente a procesos que no son estacionarios, situación habitual en las redes de ordenadores.

Siguiendo con el modelado de red, en [2] se propone el uso de distribuciones α -estables para estudiar el tráfico de red en puntos de baja agregación. Además, los autores estudian las perturbaciones que inducen ciertas anomalías en los parámetros estimados para el modelo de tráfico. Esta aproximación resulta en cierto sentido problemática, dado el alto coste computacional de este procedimiento. Observando lo expuesto en [13], esto puede ocasionar que el despliegue de esta aproximación sea inútil en ciertos escenarios. FDA incluye aproximaciones con bajo coste computacional que pueden acabar con este problema.

Otros trabajos, como [14], [15], utilizan procesos gaussianos para modelar el comportamiento de la red. Particularmente, los resultados recogidos en [14] se orientan a la planificación de la capacidad de los enlaces realizando procesos de inferencia estadística en la hora cargada, lo que lo aleja de planteamientos que se realicen dinámicamente – por analizar el comportamiento en el pico de carga. La metodología descrita en [15] se orienta a la detección de cambios sostenidos en la utilización de la red. La gaussianidad de la carga de tráfico es la base de estos y otros modelos, aunque hay numerosos estudios que no permiten asumir en general que el tráfico se comporte de este modo – véanse, por ejemplo, los resultados expuestos en [1], [2]. Muchas técnicas contempladas en FDA no tienen este problema, al no requerir hipótesis relativas a las distribuciones marginales de los parámetros considerados.

Además de los análisis comentados hasta el momento, la discriminación de comportamientos anómalos también se puede encarar desde la perspectiva de FDA. Los resultados descritos en [16], [17] se basan en el Análisis del Comportamiento de Red (*Network Behavior Analysis*, NBA). Estas soluciones utilizan detección de patrones para discriminar comportamientos anómalos que podrían indicar la presencia de algún tipo de evento en la red (*e.g.* un ataque). NBA puede ser visto desde el punto de vista funcional como una serie de funciones que describen el estado de la red, proporcionando de este modo consistencia formal al análisis y una base para el empleo de técnicas como los modelos de regresión con respuesta funcional.

Para concluir, no podemos dejar de comentar que, aunque pueda parecer contrario a la intuición, el trabajo en espacios de dimensión infinita conlleva algunas ventajas. En ese sentido, ciertas técnicas de aprendizaje automático aprovechan la simplificación de la separación de conjuntos al aumentar la dimensión para clasificar o detectar *clusters* en observaciones. Por ejemplo, las Máquinas de Vectores de Soporte (*Support Vector Machines*, SVMs) son un ejemplo bien conocido que, utilizando una inmersión de los datos en espacios de alta dimensionalidad (en algunos casos infinita), han proporcionado buenos resultados en problemas que aparecen en las actividades de gestión de red. Algunos casos de uso se exponen en [18], [19], en los que los autores exploran las ventajas de esta aproximación para la detección de anomalías

y la gestión y predicción de la Calidad de Experiencia (QoE) de los usuarios.

III. FDA COMO CONJUNTO DE HERRAMIENTAS PARA EL ANÁLISIS DE DATOS DE RED

En esta sección nos aproximaremos al procesado funcional de parámetros de red. Por cuestiones de brevedad y legibilidad, la exposición que se presenta en este trabajo no profundizará en las cuestiones y descripciones formales de los métodos considerados. Para obtener una visión más rigurosa de las opciones que plantea FDA, en [3] el autor plantea una revisión de los principales resultados en esta rama de la estadística, y en [20], [21] se exponen con mayor detalle los contenidos matemáticos resumidos en este artículo, con indicaciones acerca de implementaciones de los métodos basados en R y MatLab.

III-A. Preprocesado de datos de red

En este apartado vamos a comentar las técnicas que van a permitirnos abordar las primeras fases del tratamiento funcional de datos de red. Nuestro objetivo es plantear una línea básica de manejo de datos para adaptarlos de cara a posteriores análisis, indicando las ganancias principales derivadas de esta aproximación.

III-A.1. Representación funcional: Cuando se miden parámetros experimentalmente, no es posible obtener medidas de forma continua, sino que se obtienen una serie de valores con una cierta granularidad. Por ello, al utilizar un enfoque funcional para el análisis de datos se debe empezar por interpolar con aproximaciones que minimicen globalmente el error. Además, si el estudio a realizar lo requiere, este mecanismo de representación permite aplicar estrategias de suavizado de los datos. En la literatura de FDA, una alternativa que es comúnmente elegida para realizar este proceso de interpolación son las B-Splines, debido a sus propiedades [22]. No obstante, para obtener este tipo de representación se puede seleccionar cualquier base funcional si sus propiedades permiten el mantenimiento de la estructura de los datos de acuerdo a las necesidades del análisis posterior. En general, podemos representar el conjunto de funciones que conforman la base seleccionada con la notación $\{B_k(t)\}_{t \in \mathbb{T}, k \in \mathbb{Z}}$, siendo \mathbb{T} un intervalo real. Por otro lado, se pueden denotar las proyecciones de las observaciones funcionales con respecto a esta base como $\{\beta_k\}_{k \in \mathbb{Z}}$.

De esta forma, dada una observación $\{X_t\}_{t \in \mathbb{T}}$ podemos representarla como:

$$\{X_t\} = \sum_{j \in \mathbb{Z}} \beta_j B_j(t), \quad t \in \mathbb{T}$$

En la práctica, no es posible considerar todos los elementos de esta serie, de modo que al truncarla, aparece un término de error. Finalmente, la representación obtenida es la dada por la siguiente expresión:

$$\{X_t\} = \left[\sum_{j \in \mathbb{J}} \beta_j B_j(t) \right] + \epsilon(\mathbb{J}, \{B_j\}), \quad t \in \mathbb{T}$$

donde \mathbb{J} indica el conjunto finito de índices seleccionados y ϵ representa el término de error, que es dependiente tanto del conjunto de índices como de la base seleccionada.

Esta representación presenta diversas ventajas. En primer lugar, la cantidad de datos necesaria para representar la evolución del proceso se reduce drásticamente dado que habitualmente el número de puntos temporales de observación es mucho mayor que el número de componentes funcionales seleccionadas. Por otro lado, esta representación permite una estimación robusta de las derivadas de las trayectorias del proceso, ya que las funciones derivadas pueden calcularse analíticamente a partir de las expresiones funcionales obtenidas. Esto resulta interesante en distintos procesos analíticos: por ejemplo, al analizar comportamientos anómalos dentro de redes de comunicaciones, muchas veces no sólo es importante el valor de una cierta magnitud sino también su tasa de variación. Por otro lado, el análisis conjunto de una función y su derivada permite llevar a cabo estudios de estabilidad de sistemas dinámicos, lo que puede ser utilizado en la modelización de sistemas de red. Finalmente, como se verá a continuación, la estructura de esta representación proporciona un medio para seleccionar las componentes que contienen la información más representativa del modelo, lo que permite reducir el número de términos de la serie que representa las observaciones al utilizar una base que se ajusta de forma óptima a las características de las observaciones.

Con esta discusión, resulta claro que FDA permite obtener expresiones compactas en términos de almacenamiento de datos para representar la evolución temporal de diversos parámetros de red – *e.g. throughput* en términos de paquetes o bits por segundo, o número de *hosts* activos. Esto resulta particularmente interesante en el caso de la definición de líneas base [23], dado que se pueden obtener estimaciones definidas en tiempo continuo y con una buena aproximación al comportamiento habitual de la red.

Cabe resaltar también que estas funciones no tienen por qué tener \mathbb{R} como espacio de llegada. Si definimos el estado de la red mediante una función $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, podemos usar superficies o funciones multivariantes que permitan describir el comportamiento conjunto de un número arbitrario de parámetros. Esta capacidad es esencial en algunas tareas, como por ejemplo la detección de ciertos tipos de anomalías de red (*e.g.* ataques de Denegación de Servicio (DoS o DDoS) [8]), lo que convierte esta aproximación en una alternativa que puede ser de gran utilidad para gestores de red.

III-A2. PCA funcional: Al igual que ocurre en el caso no funcional, PCA funcional es una técnica que permite transformar la base funcional que estamos usando para representar nuestras observaciones mediante una proyección que maximiza la varianza y reduce la correlación. Conceptualmente, por tanto, resulta similar al caso no funcional. No obstante, al contrario que en el caso clásico y dado que originalmente se está trabajando con una representación en términos de una base con un sentido homogéneo, PCA funcional no conlleva una ofuscación semántica de las dimensiones, sino que por el contrario estamos obteniendo una segunda base funcional adecuada a las observaciones que estamos considerando.

No obstante, es necesario imponer algunas restricciones para evitar casos degenerados que cubran toda la varianza. Así, con el fin de no llegar a una base formada por funciones con demasiado rizado por el hecho de que cubran una gran parte de la varianza, PCA funcional requiere (i) implantar

elementos de penalización en el problema de optimización o (ii) utilizar versiones suavizadas de los datos.

La aplicación de PCA funcional en la representación de datos de red conlleva múltiples ventajas. Esta técnica selecciona combinaciones de los elementos de la base funcional previamente utilizada para representar las observaciones con un criterio basado en la proporción de varianza explicada. Por ello, se consigue un segundo nivel de compresión de los datos, manteniendo su estructura de variabilidad y obteniendo una buena aproximación en el espacio funcional donde se definen las observaciones. Por otro lado, los armónicos obtenidos describen curvas que representan aspectos de las observaciones que pueden ser interpretados en términos del orden de los autovalores asociados, frente a otras alternativas que permiten obtener versiones de los datos simplemente reducidas o filtradas.

Para completar la discusión de esta técnica, remarcamos algunas líneas que estamos evaluando en la actualidad. En primer lugar, estamos comprobando si la detección de cambios de comportamiento de los datos con respecto a alguna de las componentes puede ser utilizado como indicador de situaciones anómalas en la dinámica de la red. Este es un nuevo enfoque de la categorización de anomalías de red, basado en el estudio del comportamiento de las observaciones con respecto a las componentes principales. Finalmente, la información semántica que conlleva esta descomposición tiene un nicho de aplicación en problemas de planificación de red, en los que la definición de líneas base de comportamiento juega un papel de importancia creciente.

III-B. Análisis de datos de red

Tras haber cubierto las primeras etapas de minería de datos de red, vamos a describir algunas técnicas que aparecen en el ámbito de FDA para acometer la extracción de conclusiones a partir de las observaciones disponibles.

III-B1. Análisis basado en profundidad funcional: Las medidas de profundidad en FDA son útiles al proporcionar nociones sobre la posición relativa que ocupa una observación dentro del conjunto de observaciones. Debido a los diversos intentos realizados para definir y adaptar distintos estadísticos de orden a los datos funcionales, las medidas de profundidad se han convertido en un elemento clave para construir estadísticos que requieran inducir un cierto orden entre las observaciones.

No existe una noción única de profundidad, sino que se trata de un concepto que se define en base a ciertas propiedades de centralidad de las observaciones. En nuestro caso, vamos a seguir la definición incluida en [24], dada por la expresión

$$MS_{n,H}(x) = \min\{SL_n(x), IL_n(x)\} \quad (1)$$

donde

$$SL_n(x) = \frac{1}{n\lambda(T)} \sum_{i=1}^n \lambda\{t \in \mathbb{T} : x(t) \leq x_i(t)\}$$

$$IL_n(x) = \frac{1}{n\lambda(T)} \sum_{i=1}^n \lambda\{t \in \mathbb{T} : x(t) \geq x_i(t)\} \quad (2)$$

siendo λ la medida de Lebesgue. Con esta definición, estamos ordenando las curvas que componen la muestra usando el mínimo valor entre la proporción de “tiempo” en que se

encuentran en el hipografo ($SL_n(x)$) o epigrafo ($IL_n(x)$) de otras observaciones.

El uso de esta definición es relativamente popular, debido a que su cálculo tiene un bajo coste computacional, y a su interpretación intuitiva. Sin entrar en detalles técnicos, esta definición de profundidad funcional se basa en la proporción del dominio en el que una cierta observación domina o es dominada por el resto, lo que permite cuantificar su centralidad en el conjunto.

Por otro lado, también existen algunas definiciones multivariantes de profundidad [25], lo que permite extender los conceptos basados en la profundidad a funciones multivariantes y superficies – *e.g.* múltiples parámetros de flujos de red.

De este modo, el análisis basado en profundidad proporciona una metodología para obtener estadísticos de orden para curvas y superficies que representan un conjunto de parámetros. La definición de estos estadísticos de orden permite realizar análisis robustos para la caracterización del comportamiento típico de la red, debido al carácter aislado de los valores atípicos – de modo que producen perturbaciones mínimas de las medidas de profundidad.

La aparición de infraestructuras de red que permiten la configuración dinámica de reglas y del despliegue de recursos (*e.g.* Software Defined Networking (SDN) o la arquitecturas Application-based Network Operations (ABNO) [26]) apunta a la necesidad de establecer líneas base de comportamiento que tengan en cuenta el comportamiento de la red en todo momento. Por ello, las métricas basadas en profundidad son buenas candidatas para la definición de líneas base, dado que permiten la construcción de regiones que cubren una cierta proporción de las curvas observadas. Además, este enfoque permite caracterizar el comportamiento de la red a lo largo de todo un período (*e.g.* un día) frente a otras aproximaciones de menor dimensionalidad, que lo evalúan en términos de ciertos resúmenes estadísticos – como pueden ser medias o medianas sobre toda la actividad, o en ventanas. Las propuestas de medidas de profundidad para funciones multivariantes son también particularmente interesantes, ya que permiten considerar conjuntamente diversos parámetros. Esto es necesario al tratar ciertos eventos, como por ejemplo ataques de inundación de segmentos SYN [8].

III-B2. Homogeneidad funcional, clustering y clasificación: Cuando se tienen dos o más muestras, una cuestión que se plantea de forma natural es si ambas proceden de trayectorias del mismo proceso estocástico. En el caso de las estadísticas clásicas, existen distintos métodos que pueden ser usados para evaluar la homogeneidad de dos muestras (*e.g.* el test χ^2).

En el área de FDA, existen algunas propuestas recientes con resultados prometedores para dar respuesta a esta pregunta. En [27], los autores consideran distintas medidas de homogeneidad basadas en profundidad funcional. Para evaluar la potencia de sus métodos, utilizan una aproximación computacional, ya que no es posible (al menos en términos prácticos) el cálculo analítico de las distribuciones de los estadísticos involucrados en los contrastes de hipótesis.

El uso de la homogeneidad funcional en el área de la gestión de red proporciona un nuevo enfoque de la evaluación de la representatividad de un cierto conjunto de parámetros de red. Usando métodos como los presentados en [27], podemos

definir algoritmos automáticos para detectar conjuntos de parámetros que caractericen efectivamente el estado típico de una red – esto es, que tienen un comportamiento homogéneo en términos funcionales.

Por otro lado, una vez se han realizado análisis descriptivos y exploratorios de los datos de red, es necesario proporcionar técnicas que permitan realizar otros tipos de análisis orientados a discriminar o predecir el estado del sistema que está siendo estudiado. FDA ofrece un ecosistema completo de técnicas, que permiten realizar *clustering*, clasificación y predicción de funciones, tanto en términos de valores escalares como con modelos de respuesta funcional [3]. En este sentido, el estudio de variables aleatorias funcionales puede proporcionar un nuevo campo para la clasificación de tráfico cifrado, utilizando una aproximación basada en la determinación de los perfiles temporales de distintas clases de tráfico. Además, las técnicas de predicción con respuesta funcional permiten obtener curvas o superficies completas que modelan el comportamiento del sistema durante un período completo – *e.g.* un día. Esta aproximación no ha sido todavía evaluada, aunque sus características la posicionan como una solución que podría tener grandes ventajas en el ámbito de la planificación dinámica (*e.g.* SDNs) y otros entornos flexibles tales como las infraestructuras tipo *Cloud*.

IV. RESULTADOS EXPERIMENTALES

En esta sección vamos a presentar los resultados experimentales obtenidos aplicando algunas de las técnicas descritas en las secciones anteriores. Nuestros resultados tienen un carácter exploratorio, por lo que se centran en la visualización e interpretación de características de estas observaciones extraídas mediante algunas técnicas de FDA. Comentaremos algunas de las líneas que estamos evaluando en la actualidad, resumiendo algunos resultados alcanzados.

Si nos centramos en los estudios basados en profundidad funcional, nuestros primeros campos de aplicación se han orientado al estudio exploratorio robusto de series temporales de distintos parámetros de red, y al análisis de distribuciones de las componentes de registros de NetFlow [28].

La aplicación de profundidad funcional al análisis de series temporales muestra que este enfoque permite la identificación de fronteras y comportamientos típicos de una forma robusta. Esto es, los resultados que hemos obtenido muestran que este enfoque presenta una gran resistencia a la aparición de sucesos atípicos y de diversos regímenes de comportamiento en el seno de la red que se está analizando. Nuestros resultados también apuntan a que el análisis basado en profundidad funcional puede ser útil para la detección de días anómalos, debido a la ordenación que inducen las medidas de profundidad sobre las curvas. Finalmente, este sistema de análisis es robusto frente a cambios sostenidos en la utilización de las redes. La aparición de series temporales de baja profundidad indicaría dichos cambios, y el sistema se podría ajustar actualizando el conjunto de referencia.

En base a lo anterior, nuestros primeros estudios en términos de *clustering* y clasificación de curvas serán presentados en la siguiente sección. Estos análisis se han enfocado desde el análisis de componentes principales funcionales y el estudio del comportamiento de las observaciones respecto a dichas componentes.

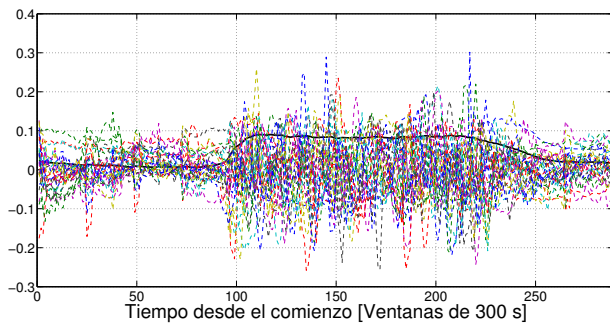


Figura 1. Comportamiento de las 30 primeras componentes principales

El conjunto de datos que se va a utilizar durante la fase de análisis se trata de un conjunto de registros de *throughput* obtenidos en la red académica de España (RedIRIS). Hemos seleccionado una serie de registros proporcionados por uno de los nodos de esta red, separando 546 días de tráfico con observaciones de *throughput* cada 300 s. Para realizar el análisis presentado, hemos utilizado la implementación de los métodos incluida en el paquete *fda* de R [29].

IV-A. Resultados del preprocesado de datos

En primer lugar, es necesario preprocesar las observaciones recabadas del nodo de comunicaciones. Para ello, seleccionamos una representación basada en B-splines de tercer grado, utilizando una base con un número de elementos similar al número de observaciones de cada curva (esto es, 288 correspondientes a los intervalos de 5 minutos que hay en un día). Una vez obtenida esta representación, aplicamos PCA funcional, analizando el comportamiento de los armónicos obtenidos. Dado que hemos utilizado PCA funcional como mecanismo de reducción de datos, hemos optado por utilizar un gran número de elementos en la base original de B-splines. Si se desea obtener un mayor grado de compresión, es posible utilizar un menor número de elementos, con la consiguiente pérdida de precisión al restringir el detalle que se puede conseguir.

El análisis de varianza explicada nos llevó a aceptar una representación utilizando una base compuesta de 30 elementos funcionales, mostrados en la Figura 1. Es interesante observar que, al no haber eliminado la componente media antes de aplicar PCA funcional, la primera componente principal, representada en negro en la figura, mantiene el patrón habitual de tráfico, ilustrado en la Figura 3.

Una vez seleccionada y fijada esta base, obtenemos los coeficientes correspondientes para las observaciones incluidas en nuestro conjunto de datos. Este comportamiento queda reflejado en la Figura 2, que muestra las densidades de los coeficientes separadas por componente principal.

Es interesante señalar que simplemente considerando el comportamiento con respecto a la primera componente principal, es posible discriminar las curvas en dos *clusters* bien diferenciados, en este caso correspondientes a días laborables y no laborables – marcados en la figura.

Por otro lado, en la Figura 3 se muestra una comparativa del comportamiento visual de las observaciones originales (parte superior) y las series recuperadas utilizando la base restringida a 30 componentes principales. Se observa que, pese a la gran reducción de información (pasamos de necesitar

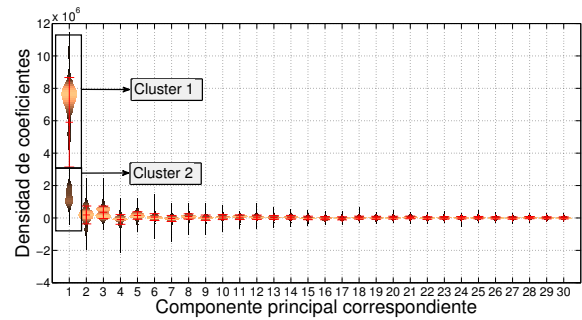


Figura 2. Densidad de los coeficientes por componente principal

288 observaciones para representar todo un día a utilizar únicamente 30), se mantiene el comportamiento de las dos series, incluso en los días que presentan un comportamiento separado de lo habitual.

IV-B. Análisis basado en profundidad

Para concluir, presentamos los resultados obtenidos al realizar un análisis basado en profundidad funcional de los datos considerados. Para obtener los resultados mostrados, restringimos en este caso la base funcional de representación a las 15 primeras componentes principales. De este modo, al suprimir componentes con baja proporción de varianza explicada se evitan perturbaciones inducidas por posibles valores atípicos. En la Figura 4 se resumen los principales resultados derivados de esta exploración, presentando curvas notables con distintos niveles de profundidad.

Para comparar el tipo de curvas que se pueden definir utilizando la noción de profundidad y su relación con las observaciones, la Figura 4 incluye en naranja claro el conjunto de observaciones con el que estamos trabajando.

En primer lugar, vamos a considerar medidas de centralidad de observaciones. Para comparar las alternativas basadas en profundidad funcional frente a otras alternativas incluimos la función de medias estimada a partir de las trayectorias del proceso estocástico que se han observado – curva de color negro sin marcadores. Junto con esta curva, representamos las curva de máxima profundidad – en cierto sentido, equivalente a la mediana. Para comparar los beneficios de restringir aún más la base funcional de representación, incluimos tanto la observación original como la curva reconstruida – en color rojo con rombos y azul con cuadrados respectivamente.

Se observa como la aparición de dos *clusters* en el conjunto de observaciones produce desviaciones en la función media, que no aporta una idea adecuada de la posición central en el conjunto de observaciones. Por otro lado, se observa también como la estimación realizada a partir de la representación funcional restringida de los datos es más estable, al suprimirse un cierto porcentaje de la varianza.

Con el fin de detectar observaciones extremas (correspondientes a días laborables y no laborables), se han evaluado también las observaciones que presentan una menor profundidad funcional, representadas en color verde con asteriscos y cruces para las curvas extremas en el epígrafe e hipógrafo de la curva más profunda, respectivamente. Por otro lado, utilizando la expresión descrita en la Ecuación (1), es posible calcular curvas que minimizan la profundidad puntualmente. En la Figura 4 se han representado también, en color naranja

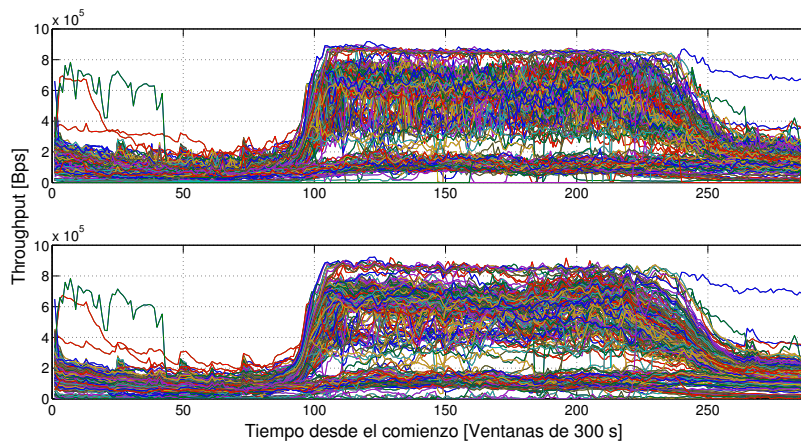


Figura 3. Comparativa entre las observaciones originales (superior) y las series reconstruidas (inferior) tras el preprocesado de los datos.

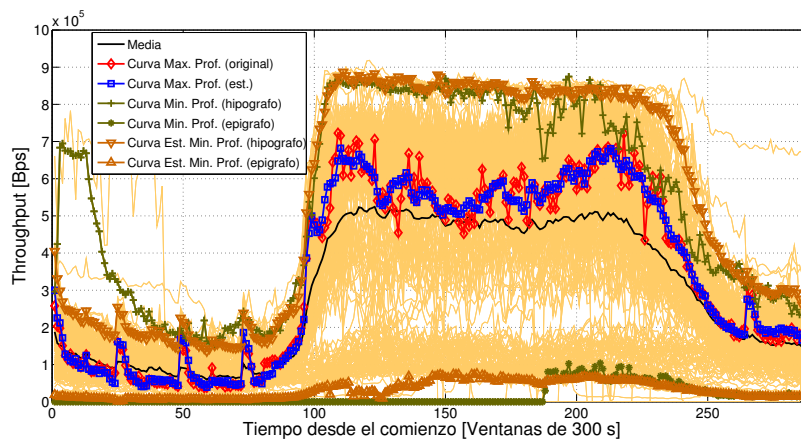


Figura 4. Resumen de los resultados obtenidos en el análisis basado en profundidad

oscuro y con triángulos, las curvas calculadas como mínimos de la profundidad, sin contabilizar el 5% de los datos más extremos para cada punto del dominio temporal. Se observa que este procesamiento permite obtener una región que cubre todas las observaciones que no presentan comportamientos atípicos.

Con estos primeros resultados, hemos mostrado que tanto el estudio de observaciones centrales como el de observaciones extremas muestran gran robustez y potencia descriptiva en términos visuales. Permiten obtener automáticamente perfiles de uso representativos, que faciliten la comparación e interpretación visual del comportamiento de la red en términos de la información disponible – como se ha visto, proporcionando márgenes que cubran un cierto porcentaje de la varianza y proporcionando medidas de centralidad de las observaciones diarias en términos globales. Estos resultados ilustran como las técnicas de FDA pueden ser aplicadas a la definición de líneas base y al establecimiento de nuevos criterios para la detección de anomalías de red.

V. CONCLUSIONES

Este trabajo se ha elaborado con el fin de enriquecer el catálogo de técnicas de análisis de datos consideradas por la comunidad de Ingeniería Telemática durante el procesado de datos y medidas procedentes de redes de comunicaciones. Para abrir nuestra discusión, hemos empezado motivando el empleo de FDA en el área de las redes de comunicaciones. A

tal efecto, hemos proporcionado una revisión de la literatura de monitorización y gestión de red, indicando aquellos puntos que podrían aprovechar las peculiaridades del análisis de datos funcionales. Esta rama de la estadística presenta técnicas que facilitan ciertas tareas del análisis de redes de comunicaciones.

Para ilustrar las ventajas que conlleva, hemos aportado una revisión de diversas técnicas de FDA relacionándolas con aplicaciones que pueden beneficiarse de ellas. Hemos estructurado nuestra exposición en torno a dos bloques que se pueden distinguir en el flujo de datos típico de las labores de gestión de red – en particular, preprocesado y análisis de los datos.

En el bloque de preprocesado, hemos comentado cómo se deben preparar los datos para poder ser tratados funcionalmente. Se han descrito los fundamentos de este análisis, mostrando la importante ganancia en términos de reducción de volumen de datos al representar los datos como observaciones funcionales. Asimismo, se ha evaluado cómo el empleo de PCA funcional permite obtener una descomposición en términos de componentes principales funcionales que abre la puerta a nuevas interpretaciones de la dinámica de la red. Hemos analizado cómo se puede vincular el comportamiento de las observaciones con respecto a las componentes principales funcionales con cambios en el patrón de uso – *e.g.* días con carga atípica, días con una aportación inusualmente alta en ciertas componentes. Además, hemos visto como esta técnica

permite visualizar *clusters* en conjuntos no homogéneos, que es otra de las acciones que FDA permite realizar.

En la parte de análisis, nos hemos centrado en la profundidad funcional. Hemos mostrado cómo al analizar la profundidad de las observaciones se puede, por un lado, encontrar el “representante” del conjunto que ocupa el lugar central (de una forma equivalente a la mediana en cierto sentido); y por otro lado, se pueden encontrar y construir curvas extremas que definen regiones cubriendo un cierto número de observaciones. Esto resulta de particular interés en tareas de planificación de red y gestión de recursos, ya que proporciona una aproximación robusta para la definición de líneas base de comportamiento.

En resumen, hemos introducido en la comunidad de Ingeniería Telemática un sistema completo de análisis que abre la puerta a una nueva comprensión de la dinámica de red. Como líneas de trabajo en desarrollo, estamos estudiando otros tipos de variables funcionales de red, la clasificación y predicción sobre datos funcionales, y el estudio de homogeneidad tanto sobre series temporales como sobre otras funciones que aparecen de forma natural en las observaciones de red – e.g. funciones de densidad. Por otro lado, también estamos estudiando el comportamiento de los residuos resultantes de la aplicación de PCA funcional, ya que podrían permitir detectar anomalías de red. FDA es una rama de la estadística que puede facilitar la realización de tareas relativas a la seguridad y prestaciones de servicios de red que, en el campo de la Estadística clásica, presentan complicaciones y limitaciones severas.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad de España (proyectos *PackTrack*, *TEC2012-33754* y *MTM2013-44045-P*).

REFERENCIAS

- [1] R. De O Schmidt, R. Sadre, N. Melnikov, J. Schönwälder, and A. Pras, “Linking network usage patterns to traffic gaussianity fit,” in *Networking Conference, 2014 IFIP*, June 2014, pp. 1–9.
- [2] F. Simmross-Wattenberg, J. Asensio-Pérez, P. Casaseca-de-la Higuera, M. Martín-Fernández, I. Dimitriadis, and C. Alberola-López, “Anomaly detection in network traffic based on statistical inference and alpha-stable modeling,” *Dependable and Secure Computing, IEEE Transactions on*, vol. 8, no. 4, pp. 494–509, July 2011.
- [3] A. Cuevas, “A partial overview of the theory of statistics with functional data,” *Journal of Statistical Planning and Inference*, vol. 147, no. 0, pp. 1 – 23, 2014.
- [4] B. Li, J. Springer, G. Bebis, and M. H. Gunes, “A survey of network flow applications,” *Journal of Network and Computer Applications*, vol. 36, no. 2, pp. 567–581, 2013.
- [5] V. Bajpai and J. Schönwälder, “A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts,” *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.
- [6] A. Papadogiannakis, M. Polychronakis, and E. P. Markatos, “Scap: Stream-oriented network traffic capture and analysis for high-speed networks,” in *Proceedings of the 2013 Conference on Internet Measurement Conference*, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 441–454. [Online]. Available: <http://doi.acm.org/10.1145/2504730.2504750>
- [7] D. Simoncelli, M. Dusi, F. Gringoli, and S. Niccolini, “Stream-monitoring with BlockMon: convergence of network measurements and data analytics platforms,” *SIGCOMM Comput. Commun. Rev.*, vol. 43, p. 29–36, 2013. [Online]. Available: <http://doi.acm.org/10.1145/2479957.2479962>
- [8] V. Moreno, P. M. Santiago del Río, J. Ramos, D. Muelas, J. L. García-Dorado, F. J. Gómez-Arribas, and J. Aracil, “Multi-granular, multi-purpose and multi-Gb/s monitoring on off-the-shelf systems,” *International Journal of Network Management*, vol. 24, no. 4, pp. 221–234, 2014. [Online]. Available: <http://dx.doi.org/10.1002/nem.1861>
- [9] J. J. Davis and A. J. Clark, “Data preprocessing for anomaly based network intrusion detection: A review,” *Computers & Security*, vol. 30, no. 6, pp. 353–375, 2011.
- [10] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. D. Kolaczyk, and N. Taft, “Structural analysis of network traffic flows,” *SIGMETRICS Perform. Eval. Rev.*, vol. 32, no. 1, pp. 61–72, Jun. 2004. [Online]. Available: <http://doi.acm.org/10.1145/1012888.1005697>
- [11] J. L. García-Dorado, J. Aracil, J. A. Hernández, and J. E. López de Vergara, “A queueing equivalent thresholding method for thinning traffic captures,” in *Network Operations and Management Symposium, 2008. NOMS 2008. IEEE*, April 2008, pp. 176–183.
- [12] K. Kyriakopoulos and D. Parish, “A live system for wavelet compression of high speed computer network measurements,” in *Passive and Active Network Measurement*, ser. Lecture Notes in Computer Science, S. Uhlig, K. Papagiannaki, and O. Bonaventure, Eds. Springer Berlin Heidelberg, 2007, vol. 4427, pp. 241–244. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-71617-4_27
- [13] K. Xu, F. Wang, and H. Wang, “Lightweight and Informative Traffic Metrics for Data Center Monitoring,” *Journal of Network and Systems Management*, vol. 20, no. 2, pp. 226–243, 2012. [Online]. Available: <http://dx.doi.org/10.1007/s10922-011-9200-6>
- [14] J. L. García-Dorado, J. A. Hernández, J. Aracil, J. E. López de Vergara, and S. López-Buedo, “Characterization of the busy-hour traffic of IP networks based on their intrinsic features,” *Computer Networks*, vol. 55, no. 9, pp. 2111 – 2125, 2011.
- [15] F. Mata, J. L. García-Dorado, and J. Aracil, “Detection of traffic changes in large-scale backbone networks: The case of the Spanish academic network,” *Computer Networks*, vol. 56, no. 2, pp. 686 – 702, 2012.
- [16] S. Saad, I. Traore, A. Ghorbani, B. Sayed, D. Zhao, W. Lu, J. Felix, and P. Hakimian, “Detecting P2P botnets through network behavior analysis and machine learning,” in *Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference on*, July 2011, pp. 174–180.
- [17] T.-E. Wei, C.-H. Mao, A. Jeng, H.-M. Lee, H.-T. Wang, and D.-J. Wu, “Android malware detection via a latent network behavior analysis,” in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, June 2012, pp. 1251–1258.
- [18] V. Menkovski, A. Oredope, A. Liotta, and A. Cuadra Sánchez, “Optimized online learning for QoE prediction,” in *Proceedings 21st Benelux Conference on Artificial Intelligence. BNAIC'09*, October 2009, pp. 169–176.
- [19] M. Nassar, O. Dabbebi, R. Badonnel, and O. Festor, “Risk management in VoIP infrastructures using support vector machines,” in *Network and Service Management (CNSM), 2010 International Conference on*, Oct 2010, pp. 48–55.
- [20] J. Ramsay and B. Silverman, *Functional Data Analysis*. 1997. Springer, New York, 1997.
- [21] J. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*. Springer New York, 2009.
- [22] P. H. Eilers and B. D. Marx, “Flexible smoothing with B-splines and penalties,” *Statistical science*, pp. 89–102, 1996.
- [23] L. H. Gibeli, G. D. Breda, R. S. Miani, B. B. Zarpelão, and L. de Souza Mendes, “Construction of baselines for VoIP traffic management on open MANs,” *International Journal of Network Management*, vol. 23, no. 2, pp. 137–153, 2013. [Online]. Available: <http://dx.doi.org/10.1002/nem.1820>
- [24] S. López-Pintado and J. Romo, “A half-region depth for functional data,” *Comput. Stat. Data Anal.*, vol. 55, no. 4, pp. 1679–1695, Apr. 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.csda.2010.10.024>
- [25] G. Claeskens, M. Hubert, L. Slaets, and K. Vakili, “Multivariate functional halfspace depth,” *Journal of the American Statistical Association*, vol. 109, no. 505, pp. 411–423, 2014.
- [26] A. Aguado, V. López, J. Marhuenda, J.-P. Fernández-Palacios *et al.*, “ABNO: a feasible SDN approach for multi-vendor IP and optical networks,” in *Optical Fiber Communication Conference*. Optical Society of America, 2014, pp. Th3I–5.
- [27] R. J. F. Díaz, R. E. Lillo, and J. Romo, “Homogeneity test for functional data based on depth measures,” *Tech. Rep.*, 2014.
- [28] R. Sommer and A. Feldmann, “Netflow: Information loss or win?” in *Proceedings of the 2Nd ACM SIGCOMM Workshop on Internet Measurement*, ser. IMW '02. New York, NY, USA: ACM, 2002, pp. 173–174. [Online]. Available: <http://doi.acm.org/10.1145/637201.637226>
- [29] M. Febrero-Bande and M. Oviedo de la Fuente, “Statistical computing in functional data analysis: the r package *fda.usc*,” *Journal of Statistical Software*, vol. 51, no. 4, pp. 1–28, 2012.

Ejecución de JavaScript en Paralelo en Secuencias de Navegación Web

Jose Losada, Juan Raposo, Alberto Pan, Paula Montoto, Manuel Álvarez.
 Departamento de Tecnologías de la Información y las Comunicaciones,
 Universidade da Coruña
 Facultad de Informática, Campus de Elviña, s/n, 15071, A Coruña (España)
 jlosada@udc.es, jrs@udc.es, apan@udc.es, pmontoto@udc.es, mad@udc.es

Resumen- Las aplicaciones de automatización Web se utilizan para diferentes propósitos, tales como vigilancia tecnológica, extracción masiva, o pruebas automatizadas de aplicaciones Web. En la mayoría de los sistemas, el componente de navegación automático se desarrolla utilizando las APIs de navegadores convencionales (ej. Internet Explorer o Firefox). Esta aproximación puede no ser la más adecuada en tareas que requieren respuestas en tiempo real. Otros sistemas utilizan navegadores desarrollados específicamente para tareas de automatización Web, y sobre ellos se pueden realizar optimizaciones basadas en ciertas peculiaridades de estos entornos. En este trabajo se presenta una novedosa técnica de optimización que permite ejecutar en paralelo código JavaScript durante la construcción de una página Web. Esta técnica se basa en un análisis inicial de las interacciones entre los scripts, para generar cierta información que será utilizada posteriormente. Los experimentos realizados sobre Webs reales muestran un alto número de scripts que pueden ser evaluados concurrentemente.

Palabras Clave- Automatización Web, JavaScript, Paralelismo.

I. INTRODUCCIÓN

La mayoría de los sitios de Internet no proporcionan interfaces programáticas para interactuar con ellos de manera automática, de forma similar a cómo lo hace una persona utilizando un navegador. Por ello, en los últimos años, ha aumentado el interés en las llamadas aplicaciones de automatización Web. Estas aplicaciones, permiten la interacción automática con los sitios Web, reproduciendo las acciones que un usuario humano ejecuta sobre un navegador y se utilizan en diferentes ámbitos, como por ejemplo, Web mashups, pruebas automatizadas de aplicaciones Web y aplicaciones de meta-búsqueda en Internet o de vigilancia tecnológica. Por ejemplo, una aplicación de automatización Web de vigilancia tecnológica dentro del ámbito de la investigación, podría ejecutarse todos los días para acceder a diferentes Webs de patentes y artículos, ejecutar de manera automática la misma búsqueda en todos ellos (ej. obtener los artículos de un área de conocimiento), y finalmente extraer los resultados y mostrarlos ordenados por algún criterio.

Una parte fundamental dentro de las tecnologías de automatización Web es la capacidad para ejecutar las llamadas secuencias de navegación Web. Una secuencia de navegación consiste en una sucesión de comandos que se ejecutan de forma automática sobre un componente de navegación especializado y donde cada uno de estos comandos representa una acción que un usuario humano podría ejecutar sobre un navegador convencional.

La Figura 1 muestra un ejemplo de una secuencia de navegación que obtiene información de patentes de la Web

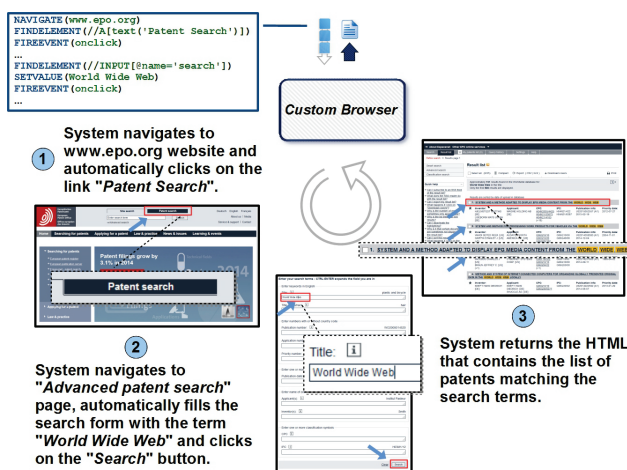


Fig. 1. Ejemplo de secuencia de navegación Web. de la Oficina Europea de Patentes (www.epo.org).

Para construir componentes de navegación, capaces de ejecutar estas secuencias, la aproximación que siguen la mayoría de los sistemas de automatización Web existentes hoy en día [1] [2] [3] [4] [5], consiste en utilizar las APIs de alto nivel de los navegadores convencionales. Esta aproximación tiene la ventaja de que evita el desarrollo de un motor de navegación desde cero y garantiza que las páginas Web a las que se accede, se comportarán de la misma manera que cuando una persona accede a ellas utilizando ese mismo navegador. El principal inconveniente de esta aproximación es que los navegadores convencionales son aplicaciones cliente (no han sido desarrollados pensando en la automatización Web) y por lo tanto, el rendimiento puede no ser el ideal en escenarios que requieren un alto número de navegadores ejecutándose de forma concurrente o cuando la tarea requiere una respuesta en tiempo real.

Otros sistemas de automatización [6] [7] [8] siguen una aproximación diferente, que consiste en desarrollar navegadores a medida (ad-hoc) especializados en la ejecución de secuencias de navegación. Estos componentes no están pensados para ser utilizados por personas y suelen ser más eficientes dado que pueden ahorrarse algunas de las tareas que realizan los navegadores convencionales (ej. renderizar el documento). No obstante, su modelo de funcionamiento a la hora de procesar los documentos HTML, suele ser muy similar al de los navegadores convencionales y además, su implementación es una tarea costosa dado que es necesario dar soporte a todas las funcionalidades de los navegadores tradicionales.

En este trabajo se presenta una novedosa técnica de optimización que permite la ejecución en paralelo del código JavaScript contenido dentro de una página Web. Esta técnica ha sido evaluada sobre un componente de navegación a medida, especializado en la ejecución de tareas de automatización Web y que da soporte a otros algoritmos de optimización presentados en trabajos previos [9].

El resto del artículo se organiza de la siguiente forma: la sección 2 describe los modelos en los que se basa esta aproximación. La sección 3 presenta una visión general del funcionamiento de los navegadores convencionales y navegadores a medida. La sección 4 explica la técnica propuesta. La sección 5 describe la evaluación experimental. La sección 6 discute los trabajos relacionados y en último lugar, la sección 7 resume las conclusiones.

II. MODELOS

A. Document Object Model (DOM)

El modelo DOM [10] describe cómo los navegadores representan las páginas Web y cómo éstas deben responder a los eventos que en ellas se generan. Para ello, una página HTML se modela como un árbol en el que cada elemento se representa con un tipo de nodo apropiado. Un tipo de nodos muy importante son los nodos script, utilizados para almacenar contenido ejecutable, típicamente, a través de la utilización del lenguaje interpretado JavaScript.

Todos los nodos contenidos en el árbol DOM pueden recibir eventos producidos (de forma directa o indirecta) por acciones del usuario. Los diferentes tipos de eventos existen para acciones como: pulsar sobre un elemento de la página (*click*), indicar que una página se ha cargado (*load*), etc.

Cada nodo puede registrar un conjunto de manejadores para procesar los diferentes tipos de eventos. Un manejador puede ejecutar código JavaScript arbitrario y puede acceder al árbol DOM para realizar acciones como crear, modificar o eliminar nodos, o disparar nuevos eventos. El ciclo de vida para el procesamiento de un evento consta de varias fases en las que el evento puede ser gestionado, tanto en el nodo de destino, como en cualquiera de sus ancestros.

B. Dependencias entre nodos del árbol DOM

En trabajos previos de los autores [9], se introdujo el concepto de dependencias entre nodos del árbol DOM. Este concepto es clave para el desarrollo de la técnica de optimización que permite ejecutar JavaScript en paralelo y se puede resumir con las siguientes definiciones:

Definición 1: se dice que existe una dependencia entre dos nodos $n1$ y $n2$ cuando el nodo $n2$ es necesario para la correcta ejecución del nodo $n1$. Se dice que el nodo $n2$ es una dependencia del nodo $n1$ y se denota $n1 \rightarrow n2$. Las siguientes reglas definen este tipo de dependencias:

a) Si el código script de un nodo $s1$ utiliza un elemento declarado en un nodo script $s2$ (ej. una función o una variable), entonces $s1 \rightarrow s2$. Justificación: para poder ejecutar el código de script del nodo $s1$, debe ejecutarse con anterioridad el nodo $s2$.

b) Si el código script de un nodo s utiliza un nodo n , entonces $s \rightarrow n$. Justificación: para poder ejecutar el código script del nodo s , debe cargarse con anterioridad el nodo n . Por ejemplo, si s obtiene una referencia al nodo A utilizando

la función *getElementById* del objeto *document* y navega a la URL especificada en su atributo *href*, entonces no se podrá ejecutar de forma correcta el código de script de s , si no se carga con anterioridad el nodo A .

c) Si el código script de un nodo s realiza una modificación en un nodo n , entonces $n \rightarrow s$. Justificación: la acción realizada por s puede ser necesaria en una acción posterior sobre el nodo n . Por ejemplo, si s modifica el atributo *action* de un nodo *form* para establecer la URL destino, entonces no se podrá realizar el envío del formulario a no ser que antes s se ejecute.

Definición 2: se dice que existe una dependencia condicionada a que el evento e se dispare sobre el nodo n , entre dos nodos $n1$ y $n2$, cuando el nodo $n2$ es necesario para la correcta ejecución del nodo $n1$, cuando se dispara el evento e sobre el nodo n . Se denota como $n1 \rightarrow^{e|n} n2$. Este tipo de dependencias se definen mediante un conjunto de reglas análogas a las explicadas en la definición 1, con la diferencia de que en este caso, involucran nodos que contienen manejadores de eventos:

a) Si el código script de un manejador de eventos l para el evento e sobre el nodo n utiliza un elemento declarado en un nodo script s (ej. una función o una variable), entonces $n \rightarrow^{e|n} s$. Justificación: si el evento e se emite sobre el nodo n , entonces se ejecutará el manejador de eventos l , el cual requiere que se ejecute antes el nodo script s .

b) Si el código script de un manejador de eventos l , para el evento e , sobre el nodo $n1$, utiliza un nodo $n2$, entonces $n1 \rightarrow^{e|n1} n2$. Justificación: si se emite el evento e sobre $n1$, entonces se ejecutará el manejador de evento l y por tanto, se deberá cargar con anterioridad el nodo $n2$.

c) Si el código script de un manejador de eventos l para el evento e sobre el nodo $n1$ realiza una modificación en un nodo $n2$, entonces $n2 \rightarrow^{e|n1} n1$. Justificación: la acción realizada por l puede ser necesaria para que $n2$ sea utilizado con posterioridad. Por ejemplo, si l modifica el atributo *action* de un nodo *form* para establecer la URL destino, entonces no será posible enviar el formulario a no ser que se ejecute con anterioridad l . Debido a que l sólo se ejecutará cuando se dispare el evento e sobre $n1$, entonces $n1$ es necesario.

De manera adicional, para calcular el conjunto completo de dependencias entre los nodos del árbol DOM, es necesario aplicar las siguientes propiedades de transitividad (todas ellas están detalladas en [9]):

Propiedad 1: si $n1 \rightarrow n2$ y $n2 \rightarrow n3$ entonces $n1 \rightarrow n3$.

Propiedad 2: si $n1 \rightarrow^{e|n} n2$ y $n2 \rightarrow n3$ entonces $n1 \rightarrow^{e|n} n3$.

Propiedad 3: si $n1 \rightarrow n2$, y $n3 \rightarrow n2$ porque $n2$ es un nodo script que realiza una modificación en $n3$, entonces $n3 \rightarrow n1$.

Propiedad 4: si $n1 \rightarrow^{e|n} n2$ y $n3 \rightarrow n2$ porque $n2$ es un nodo script que realiza una modificación en $n3$, entonces $n3 \rightarrow^{e|n} n1$.

La Figura 2 ilustra un ejemplo de dependencias entre nodos. En este ejemplo, el script con identificador $s1$ define la función f . Esta función accede al nodo *link* con identificador $l1$ (utilizando la función *getElementById* del objeto *document*), por lo tanto $s1 \rightarrow l1$. El nodo script con identificador $s2$ utiliza la función f , por lo tanto $s2 \rightarrow s1$. Y aplicando la primera regla de transitividad $s2 \rightarrow l1$. Por lo tanto $l1$ y $s1$ son dependencias de $s2$ (ambos son necesarios para su correcta ejecución).

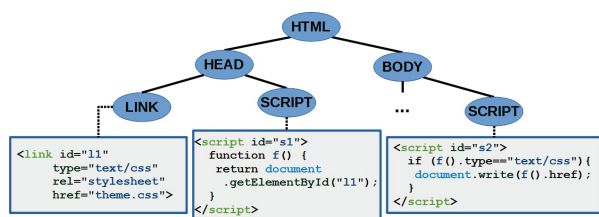


Fig. 2. Ejemplo de dependencias entre nodos.

III. VISIÓN GENERAL

En esta sección se detalla la arquitectura y el modelo de funcionamiento de los navegadores convencionales (sección 3.A) y de los navegadores a medida (sección 3.B).

A. Navegadores Web Convencionales

Un navegador Web es un software desarrollado para obtener y presentar contenidos descargados de Internet, sobre todo documentos HTML. La Figura 3 muestra la arquitectura de referencia de los navegadores convencionales [11] con sus principales componentes de alto nivel: interfaz de usuario, motor de navegación (*browser engine*) y motor de renderización (*rendering engine*). Estos componentes de alto nivel hacen uso de los subsistemas auxiliares que también se muestran en la figura: subsistema de red, subsistema de persistencia, intérprete de JavaScript, procesador de HTML y subsistema de visualización (*display backend*).

El motor de navegación es la interfaz para manipular el motor de renderización. Este componente proporciona un API para invocar acciones de alto nivel del navegador (ej. iniciar la carga de una URL, ir a la página anterior o recargar la página actual).

El motor de renderización representa el núcleo central de los navegadores Web al ser el responsable de procesar y mostrar el contenido HTML.

Durante el procesamiento de los documentos HTML, se disparan una serie de eventos en cascada, la mayoría de los cuales se procesan de manera secuencial dentro del motor de renderización. La Figura 4 muestra estas etapas:

1. Descarga y decodificación: el flujo de datos se descarga del sitio Web y se descomprime. Los navegadores utilizan cachés para evitar descargas innecesarias.

2. Construcción del árbol DOM: por motivos de eficiencia este proceso se ejecuta de manera incremental (por ejemplo utilizando un procesador SAX [12]) y a medida que se van descubriendo elementos, éstos se procesan (ej. ficheros JavaScript). Los navegadores tienen la capacidad de realizar múltiples descargas en paralelo y además, aplican distintos mecanismos de optimización durante esta fase (ej. el procesamiento especulativo de *Mozilla Firefox* [13]).

3. Aplicación de CSS: las hojas de estilo contienen información para visualizar los elementos que se muestran en la pantalla. Cada hoja de estilo contiene una serie de reglas que se aplicarán sobre los nodos del DOM.

4. Generación del esquema de diseño (*layout*) y renderización: el diseño de la página contiene rectángulos con atributos visuales (ej. dimensiones y colores). El proceso de renderizado dibuja este diseño en la pantalla del navegador utilizando las primitivas de visualización. Durante la ejecución de JavaScript también se interactúa con el diseño de la página y éste puede verse modificado dinámicamente.

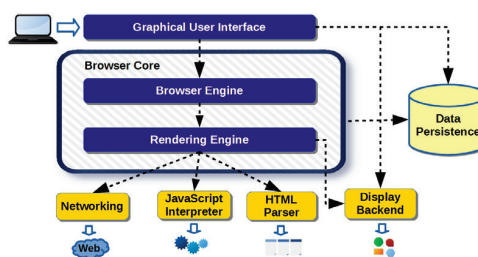


Fig. 3. Arquitectura de referencia de los navegadores convencionales.

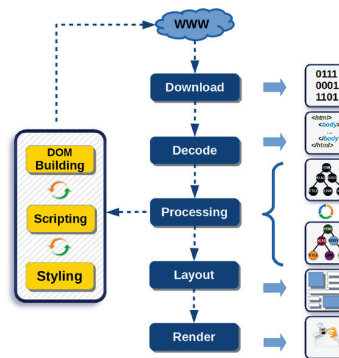


Fig. 4. Etapas de procesamiento en el motor de renderización.

Por la semántica del lenguaje JavaScript, los navegadores ejecutan de manera secuencial los scripts contenidos en un documento HTML, salvo dos excepciones:

1) Los scripts que contienen el atributo *async* (forma parte de la especificación de HTML5 [14]) se pueden ejecutar de manera asíncrona sin esperar a que otros scripts finalicen su ejecución.

2) Los objetos *Web Worker* (también forman parte de la especificación de HTML5 [14]) permiten la ejecución de JavaScript en segundo plano. Su limitación más importante es que no pueden acceder a los objetos del árbol DOM.

La Figura 5 ilustra un ejemplo real del navegador *Google Chrome* cargando una página Web. La figura muestra los distintos eventos que se disparan en este navegador: procesamiento de HTML (*Parse HTML*), descarga de objetos (*Send Request*), ejecución de JavaScript (*Evaluate Script* y *Function Call*), formateado del esquema de diseño (*Layout* y *Recalculate Style*) y dibujado en la pantalla (*Paint* y *Composite Layers*).

Durante el procesamiento del flujo de datos HTML, a medida que se van descubriendo nuevos recursos, se van disparando las acciones correspondientes. La página Web contiene hojas de estilo, imágenes y scripts. Esos elementos se descargan en paralelo (eventos *Send Request x 2* y *Send Request x 9*).

La evaluación del código JavaScript se realiza sobre todo de manera secuencial aunque existe una pequeña paralelización porque algunos scripts contienen el atributo *async* (*Function Call x 4* se ejecuta en paralelo con la evaluación del script *github.js*). Los eventos *Function Call x 4* y *Function Call x 9* representan, sobre todo, la ejecución de código generado con la función *SetTimeout*. En la figura aparece un cuarto script (*api.js*) que se genera de forma dinámica y que contiene el atributo *async* con lo que también podría ser evaluado en segundo plano. Los eventos de renderización y de dibujado en la ventana del navegador se generan durante todo el proceso de carga.

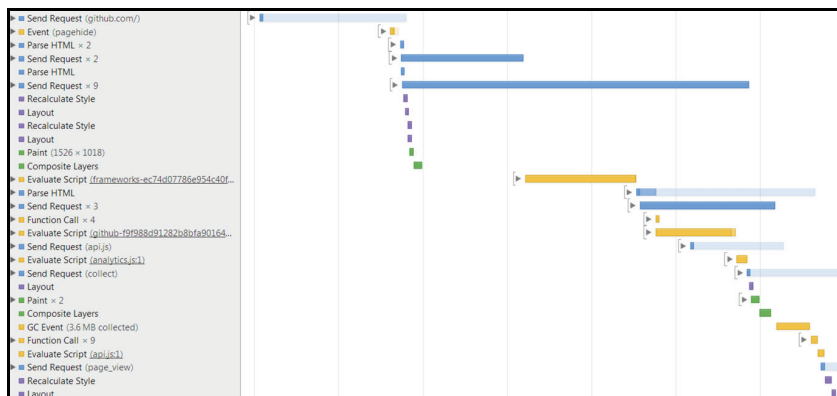


Fig. 5. Eventos en Google Chrome cargando una página Web.

B. Navegadores Web a medida

Los navegadores ad-hoc son componentes de navegación que utilizan como base la arquitectura de los navegadores convencionales pero están desarrollados de forma específica para la ejecución de secuencias de navegación en los entornos de automatización Web. Estos componentes suelen emular el comportamiento de los navegadores tradicionales y están desarrollados con dos objetivos principales. El primero consiste en alcanzar una simulación del navegador convencional lo más precisa posible. Si esto no se consigue, es posible que la ejecución de la secuencia dé lugar a una página Web incorrecta. El segundo objetivo consiste en alcanzar un alto nivel de eficiencia que permita dar soporte a los tiempos de respuesta requeridos en algunas tareas de automatización Web.

La Figura 6 muestra la arquitectura típica de los navegadores a medida. Una de las principales diferencias con los navegadores convencionales es que los navegadores a medida no necesitan interfaz de usuario. Esto aumentará la eficiencia de estos sistemas porque se evita la construcción del esquema de diseño con la información de visualización y se evita también la renderización del documento.

En los navegadores ad-hoc, el motor de navegación es también el punto de entrada para acceder al motor de renderización, pero en este caso las acciones a ejecutar no provienen de la interfaz de usuario sino que provienen de la secuencia de navegación.

En los entornos de automatización Web, las secuencias de navegación son conocidas ‘a priori’ y además, una misma secuencia se ejecuta múltiples veces. Esta peculiaridad se puede utilizar para extraer cierta información útil (durante una primera ejecución de la secuencia) con el objetivo de utilizarla en las siguientes ejecuciones y mejorar la eficiencia.

La técnica de optimización propuesta en este trabajo explota esta peculiaridad de tal manera que durante una primera ejecución de la secuencia, se analizan las dependencias entre los scripts contenidos en la página Web. Esta información se almacena y se utiliza en las siguientes ejecuciones de la misma secuencia para ejecutar en paralelo aquellos scripts identificados como independientes entre sí.

Para ello, el componente de navegación trabajará en dos fases diferentes: optimización y ejecución. Durante la fase de optimización se calculará la información necesaria para poder ejecutar los scripts en paralelo. Más en detalle, se calculará un grafo de dependencias entre scripts en el cual, para cada script S contenido en la página, se identificarán todos los

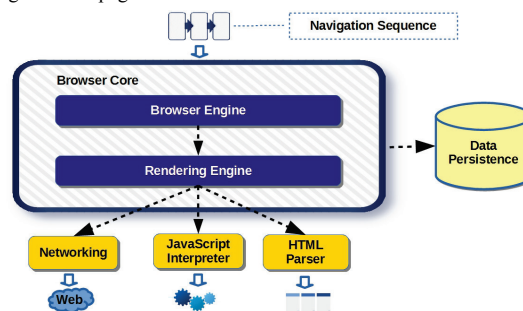


Fig. 6. Arquitectura de los navegadores a medida.

demás scripts que deben ejecutarse con anterioridad para que S funcione de forma correcta. Durante la segunda fase (fase de ejecución), se utilizará el grafo de dependencias para ejecutar en paralelo los scripts que no tienen dependencias entre sí (durante el proceso de carga del documento HTML).

IV. EJECUCIÓN EN PARALELO DE JAVASCRIPT

Los navegadores convencionales realizan la evaluación del JavaScript de manera secuencial excepto las dos excepciones comentadas en la sección 3.A.

En este apartado se detalla la técnica de optimización que permite la ejecución en paralelo de aquellos scripts de una misma página que no tienen dependencias entre sí. Un navegador a medida que utilice esta técnica, debe trabajar en dos fases: optimización y ejecución. Este proceso se detalla en [9] y se puede resumir de la siguiente manera:

En la fase de optimización, la secuencia de navegación se ejecuta una vez, se monitoriza la evaluación del código JavaScript y se obtienen las dependencias entre los nodos del árbol DOM (aplicando las reglas definidas en la sección 2.B). Estas dependencias se representan en un grafo (este grafo se utiliza en el trabajo previo [9] para diseñar otras técnicas de optimización). Con este grafo se puede construir una estructura adicional que sólo contendrá las dependencias entre los nodos de tipo script y que permitirá desarrollar la técnica de optimización propuesta. Esta nueva estructura se denominará grafo de dependencias entre scripts y contendrá, para cada script S definido en la página, una lista con los otros scripts que se deben ejecutar con anterioridad para que S funcione de manera correcta (ej. esos scripts definen variables utilizadas en S).

El Algoritmo 1 describe el pseudo-código para la generación del grafo con las dependencias entre los scripts presentes en el documento. La función principal (*BuildScriptDependencyGraph*) tiene dos argumentos: el

```

Algorithm: builds the scripts dependency graph.
Inputs: 1. dependencyGraph: graph with the dependencies between DOM nodes.
          2. DOM: DOM tree of the document.
Output: scriptDependencyGraph with the dependencies between script nodes.

1: function BuildScriptsDependencyGraph(dependencyGraph, DOM) {
2:   var scriptDependencyGraph = new Graph()
3:   for each node in DOM {
4:     if (node.isScript()) {
5:       AddScriptDependencies(node, dependencyGraph, node, scriptDependencyGraph);
6:       if (not scriptDependencyGraph.Contains(node)) {
7:         scriptDependencyGraph.AddSafeScript(node)
8:       }
9:     }
10:  }
11:  return scriptDependencyGraph
12: }

Algorithm: recursively collect script dependencies starting in the node dependencies.
Inputs: 1. node: node to check if contains script dependencies
          2. dependencyGraph: graph with the dependencies between DOM nodes.
          3. script: script node to add dependencies.
          4. scriptDependencyGraph: io parameter with the dependencies between scripts.
Output: script dependencies will be added to scriptDependencyGraph.

1: function AddScriptDependencies(node, dependencyGraph, script, scriptDependencyGraph) {
2:   var nodeDependencies = dependencyGraph.Get(node);
3:   for each nodeDependency in nodeDependencies {
4:     if (nodeDependency.isScript()) {
5:       scriptDependencyGraph.Add(script, nodeDependency)
6:     }
7:   }
8:   AddScriptDependencies(nodeDependency, dependencyGraph, script, scriptDependencyGraph)
9: }
    
```

Alg. 1. Construcción del grafo de dependencias entre scripts.

```

Algorithm: tests if a script is ready to run.
Inputs: 1. script: script node to test if it is ready to run.
          2. scriptDependencyGraph: graph with dependencies between scripts.
Output: true if the script can be executed.

1: function IsReadyForExecution(script, scriptDependencyGraph) {
2:   if (scriptDependencyGraph.IsSafeScript(script)) {
3:     return true
4:   }
5:   var dependencies = scriptDependencyGraph.Get(script);
6:   for each scriptDependency in dependencies {
7:     if (not scriptDependency.HasFinished()) {
8:       return false
9:     }
10:  }
11:  return true
12: }
    
```

Alg. 2. Detección de scripts preparados para ejecutarse.

grafo con las dependencias entre los nodos del DOM (*dependencyGraph*), calculado durante la ejecución de la secuencia (este grafo contiene dependencias directas pero no dependencias transitivas), y el árbol DOM de la página. El algoritmo itera por todos los nodos del árbol para localizar aquellos que son de tipo script y cada vez que se identifica a uno, se realiza una búsqueda dentro del grafo de dependencias, para identificar otros scripts que sean dependencias de éste (utilizando la función auxiliar *AddScriptDependencies*). Por motivos de optimización que se explicarán más adelante, los scripts que no dependan de ningún otro script, se almacenan dentro de una lista especial (línea 7). Esta lista se denominará lista de scripts seguros.

La Figura 7 muestra un ejemplo de generación de un grafo de dependencias entre scripts sobre un árbol que contiene, entre otros nodos, cuatro scripts. El primer script, *s1*, define la variable *x* y no tiene dependencias con los otros scripts de la página. El segundo script, *s2*, define la variable *y*, y tampoco tiene dependencias con otros scripts (por lo tanto, su ejecución se podrá paralelizar con la de *s1*). El tercer script, *s3*, utiliza la variable *y*, definida con anterioridad en *s2*. Por lo tanto, el grafo de dependencias entre scripts incluirá una dependencia *s3*→*s2*, lo que significa que *s3* no puede empezar su ejecución hasta que *s2* haya finalizado. Por otro lado, *s3* no tiene ninguna dependencia con el primer script *s1*, con lo que la evaluación de éstos se podrá realizar de manera concurrente. Por último, el cuarto script, *s4*, utiliza tanto la variable *x* como la variable *y*. Por lo tanto, el grafo de dependencias entre scripts incluirá una dependencia de *s4*→*s1* y de *s4*→*s2*. Esto implica que *s4* no podrá paralelizarse ni con *s1* ni con *s2* (pero sí con *s3*). La Figura 8 muestra otro ejemplo de generación de grafo que,

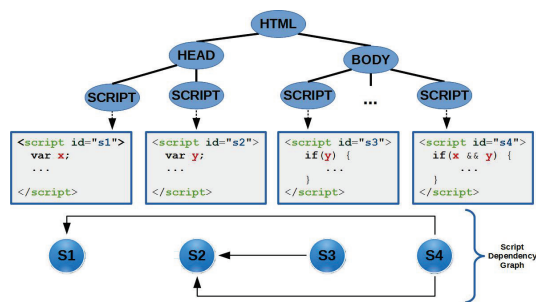


Fig. 7. Ejemplo de grafo de dependencias entre scripts.

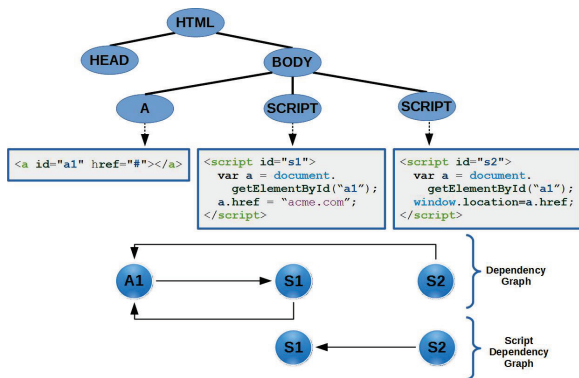


Fig. 8. Ejemplo de grafo de dependencias entre scripts con nodos.

en este caso, involucra a nodos de tipo script y también a otro tipo de nodos del DOM. En este ejemplo se incluyen dos nodos script, identificados como *s1* y *s2* y un nodo de tipo enlace, *a1*. El script *s1*, accede a *a1* a través de la función *getElementById*. Aplicando las reglas para el cálculo de dependencias, *s1*→*a1*. Pero además, *s1* también modifica el atributo *href* de *a1*. Por lo tanto *a1*→*s1*. El script *s2*, también accede a *a1*, con lo que *s2*→*a1*. Además, *s2* utiliza el atributo *href* de *a1*. Este atributo ha sido modificado con anterioridad, durante la ejecución de *s1*. Por lo tanto, en este caso, para la correcta ejecución del script *s2*, *s1* se debe ejecutar con anterioridad y no puede existir paralelización entre ellos. Esta situación se refleja en el grafo de dependencias entre scripts, con una dependencia *s2*→*s1*.

Una vez generado el grafo de dependencias entre scripts, es necesario utilizar algún mecanismo que permita identificar a los nodos contenidos en él (durante la fase de ejecución) y para ello, se utilizarán expresiones XPath [15], generadas con un algoritmo detallado en [9]. A continuación, durante la fase de ejecución, se utiliza este grafo para detectar aquellos scripts que se pueden evaluar al mismo tiempo. Esta fase involucra las siguientes ejecuciones de la misma secuencia.

El Algoritmo 2 muestra el pseudo-código de la función que determina si un script está preparado para ejecutarse. En primer lugar, el algoritmo comprueba si es un script seguro (*IsSafeScript*, línea 2), esto es, un script que no tiene dependencias con ningún otro script de la página y por lo tanto, durante la fase de optimización, ha sido almacenado en la lista de scripts seguros. En caso afirmativo, no se realiza ninguna comprobación adicional y el algoritmo determina que el script está preparado para ejecutarse. Si se trata de un script que tiene dependencias, se comprueba si éstas han finalizado (utilizando la función *HasFinished*, línea 7).


```

Algorithm: executes one pending event in the input thread.
Inputs: 1. thread: idle thread ready to execute pending events.
        2. eventQueue: queue containing pending events.
        3. scriptDependencyGraph: graph with script dependencies.
Output: pending event is dispatched in the input thread.

1: function execute(thread, eventQueue, scriptDependencyGraph) {
2:   if (eventQueue.isEmpty()) {
3:     return
4:   }
5:   if (thread.IsMainThread()) {
6:     thread.dispatch(eventQueue.Pop())
7:   } else {
8:     for each event in eventQueue {
9:       if (not event.IsScriptEvent()) {
10:        continue
11:      }
12:      var script = event.GetScript()
13:      if (script.IsAsync() or
14:         IsReadyForExecution(script, scriptDependencyGraph)) {
15:        thread.dispatch(event)
16:      }
17:    }
18:  }
19: }
    
```

Alg. 3. Ejecución de scripts.

Cuando todas las dependencias de un script han finalizado, el algoritmo determina que el script está listo para ejecutarse.

El Algoritmo 3 muestra el proceso de selección de scripts para su ejecución (así como la selección de otras acciones de navegación que se generan durante la ejecución de la secuencia). Este algoritmo está diseñado para un navegador ad-hoc que contiene los siguientes componentes en su motor de renderización:

1. Cola de eventos: almacenará diferentes tipos de acciones para su ejecución, como por ejemplo, procesado de código HTML, ejecución de scripts, ejecución de manejadores de acciones de usuario, etc. La ejecución de eventos puede producir nuevos eventos que se almacenarán también en la cola.

2. Thread principal: ejecuta cualquier evento almacenado en la cola de eventos (ejecución de scripts, procesado de HTML, etc.). Si la secuencia de navegación se ejecuta sin hacer uso de la técnica de ejecución de scripts en paralelo, todos los eventos se ejecutan de forma secuencial en este thread, emulando el comportamiento síncrono de un navegador convencional.

3. Pool de threads: sólo ejecutan eventos de evaluación de scripts en paralelo. Este pool se utiliza sólo si el grafo de dependencias entre scripts está disponible para la página que se está cargando. En caso contrario, los scripts se ejecutarán de forma secuencial en el thread principal.

Este algoritmo recibe tres parámetros: el thread que va a ejecutar el evento, la cola que almacena los eventos pendientes de ejecución y el grafo con la información de las dependencias entre los scripts. En primer lugar, el algoritmo comprueba si la cola de eventos pendientes contiene algún evento (*IsEmpty*, línea 2) y se detiene en caso de que la cola esté vacía. A continuación, se comprueba si el thread llamante es el thread principal. En caso afirmativo, este proceso ejecutará el primer evento en la cola de eventos (función *Pop*, línea 6). El thread principal ejecuta eventos de cualquier tipo, incluidos eventos de evaluación de scripts. Si el thread llamante es un thread del pool, el algoritmo itera por los eventos almacenados en la cola, buscando eventos de ejecución de JavaScript (*IsScriptEvent*, línea 9). El pool de threads sólo ejecutará este tipo de eventos. Cuando aparece un evento de evaluación JavaScript, se comprueba si todas las dependencias del nodo asociado a ese evento han finalizado. Un script se considerará preparado para su ejecución cuando contiene el atributo *async* (*IsAsync*, línea 13) o cuando no tiene dependencias pendientes de

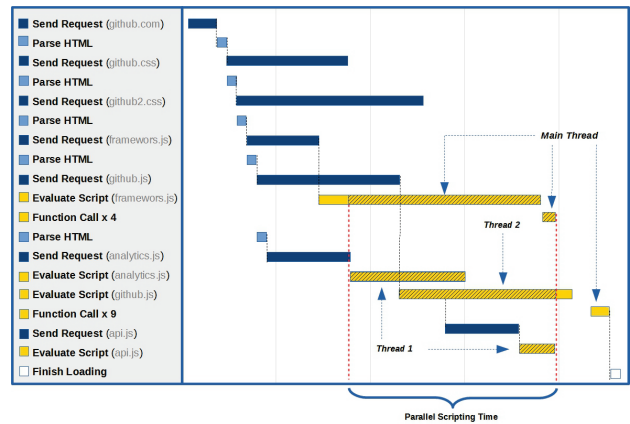


Fig. 9. Eventos generados en el navegador a medida.

finalización (*IsReadyForExecution*, línea 14, algoritmo detallado con anterioridad).

Cuando un thread termina la ejecución de un evento, vuelve a invocar a esta función para obtener y ejecutar un nuevo evento almacenado en la cola. Cada vez que un script finaliza su ejecución (tanto los ejecutados en el thread principal, como los ejecutados en el pool), el grafo de dependencias se actualiza marcándolo como finalizado. Una vez hecho esto, otros scripts pueden modificar su estado y pasar a estar listos para ejecutarse.

Puede suceder que un cambio en el código fuente provoque que algunos nodos script no puedan ser identificados en la nueva versión del documento. Cuando esto sucede, los scripts con dependencias sin identificar no se podrán paralelizar y serán ejecutados en el thread principal tras finalizar los otros eventos almacenados en la cola, siguiendo el orden natural. Si en la página se detectan nuevos scripts, el grafo de dependencias entre scripts se invalida y debe ser recalculado.

La Figura 9 muestra una simulación, mediante un cronograma, de los eventos generados en el navegador ad-hoc cuando utiliza la técnica de evaluación de scripts en paralelo. Esta figura muestra el proceso de carga de la misma página Web que se mostraba en la Figura 5 en el cronograma de eventos de *Google Chrome*. Los scripts contenidos en la página han sido identificados como scripts seguros (sin dependencias entre ellos) y por lo tanto su ejecución se podrá realizar en paralelo. Las llamadas de red se representan con las acciones "Send Request", los eventos de procesado de HTML se representan con las acciones "Parse HTML" y los eventos de evaluación de JavaScript con las acciones "Evaluate Script" y "Function Call". El proceso de carga involucra la descarga de ficheros CSS y JavaScript (uno de ellos generado de forma dinámica). Durante el procesado del código fuente HTML, el analizador descubre los objetos externos de manera incremental e inicia de forma automática su descarga en paralelo. Además, los scripts también se evaluarán en paralelo, dado que durante la fase de optimización no se han detectado dependencias entre ellos. Incluso el orden en el que se ejecutan puede ser diferente del orden natural. Cada uno de estos scripts se puede ejecutar en un thread diferente. En la figura, se muestran con líneas de trazos discontinuos los eventos de ejecución de scripts que se han ejecutado en paralelo. Los eventos "Function Call x 4" y "Function Call x 9", se han generado dinámicamente durante la evaluación de *frameworks.js* (para determinar si estos

eventos se pueden paralelizar, se utilizan las dependencias de este script). “Function Call x 4” se ejecuta en paralelo con el script *github.js*. “Function Call x 9”, comienza su ejecución tras un intervalo de tiempo (este evento ha sido generado con la función *setTimeout*). El script *api.js* ha sido generado de manera dinámica (utilizando la función *write* del objeto *document*) y además, también contiene el atributo *async*.

V. EVALUACIÓN

Para evaluar la técnica de optimización propuesta, se ha utilizado un navegador ad-hoc que simula el comportamiento de *Microsoft Internet Explorer*. Los algoritmos detallados en el capítulo anterior, se han implementado dentro del motor de renderización de este navegador.

Para realizar esta evaluación experimental, se han seleccionado diversos sitios Web incluidos en la lista de sitios más visitados según el ranking publicado por Alexa [16] y se ha ejecutado una secuencia de navegación que carga una página representativa de cada uno de estos sitios. La máquina seleccionada para las pruebas tiene 4 procesadores y 16GB de memoria RAM.

En primer lugar, se ha realizado una ejecución para obtener el grafo de dependencias entre scripts. A continuación se ha realizado una ejecución normal de la secuencia (evaluando los scripts de manera secuencial) y en último lugar, una ejecución optimizada (evaluando los scripts en paralelo utilizando el grafo de dependencias calculado con anterioridad). Para prevenir problemas e inconsistencias, derivadas de pequeños cambios en la misma página cuando se accede a ella en dos momentos diferentes, se han realizado 10 ejecuciones diferentes de cada secuencia de navegación y los datos que se muestran son la media de esas 10 ejecuciones.

La Tabla 1 muestra el tiempo que ha empleado el navegador ad-hoc en la ejecución de JavaScript (en milisegundos) tanto en la ejecución normal (primera columna) como en la ejecución optimizada (segunda columna). Como se puede observar, la mejora que se produce en el tiempo medio de ejecución cuando se utilizan las técnicas de evaluación de scripts en paralelo es de un 22% (este porcentaje varía desde el 6% en el peor de los casos hasta el 48% en el mejor de los escenarios). Descartando los resultados que quedan fuera del rango que comprende la media ± la desviación típica, la mejora se sitúa en un 20% y la mediana se sitúa también en el 20%. En la tabla no se detalla el tiempo empleado en acceder al grafo de dependencias entre scripts dado que en la mayoría de las ocasiones, este tiempo no llega a 1 milésima de segundo.

La Tabla 2 muestra un análisis de los scripts ejecutados en cada una de las secuencias. En esta tabla se muestra el número total de scripts evaluados, y de estos, se muestra cuántos han sido ejecutados en los threads del pool destinados a la evaluación en paralelo. En la tabla también se muestra el número de scripts detectados como completamente seguros, esto es, sin ninguna dependencia con otros scripts de la misma página. Como se puede observar, el número de scripts evaluados en los threads del pool varía desde el 14% en el peor de los escenarios hasta el 80% en el mejor de ellos. La media de scripts ejecutados en threads del pool es del 45%, la mediana es el 44% y la media descartando los valores que quedan fuera del rango de la media ± la

Tabla 1
TIEMPO DE EJECUCIÓN DE JAVASCRIPT EN MILLISEGUNDOS

	NORMAL	PARALELO		NORMAL	PARALELO
360.CN	392	304 (23%)	MEDIAFIRE.COM	455	357 (22%)
ALIBABA.COM	97	81 (17%)	PETFLOW.COM	727	557 (24%)
ALLEGRO.PL	425	348 (19%)	PINTEREST.COM	210	173 (18%)
AMAZONWS.COM	319	234 (27%)	REDIFF.COM	290	245 (16%)
BBC.COM	203	144 (30%)	REUTERS.COM	831	725 (13%)
BET365.COM	195	116 (41%)	RT.COM	605	478 (21%)
BILD.DE	1098	806 (27%)	SCRIPBD.COM	689	579 (16%)
BLOGGER.COM	99	65 (35%)	SOFTONIC.COM	537	483 (11%)
BLOGSPOT.COM	112	66 (42%)	SOURCEFORGE.NET	504	327 (36%)
BLOOMBERG.COM	632	529 (17%)	SPEEDTEST.NET	770	688 (11%)
BOOKING.COM	518	438 (16%)	STACKEXCHANGE.COM	255	222 (13%)
CNET.COM	235	174 (26%)	TAOBAO.COM	191	153 (20%)
ENGADGET.COM	447	274 (39%)	TARINGA.NET	1494	1405 (6%)
FORBES.COM	432	390 (10%)	TECHCRUNCH.COM	249	197 (21%)
GITHUB.COM	330	237 (29%)	THEFREEDICTIONARY.COM	503	424 (16%)
GIZMODO.COM	80	60 (25%)	TIME.COM	329	204 (38%)
GSMARENA.COM	390	335 (15%)	TRIPADVISOR.COM	131	103 (22%)
IGN.COM	1228	1036 (18%)	TUMBLR.COM	546	418 (24%)
IKEA.COM	81	69 (15%)	UPLOADED.NET	412	344 (17%)
IMGUR.COM	686	549 (20%)	UPS.COM	1167	1013 (14%)
INDIATIMES.COM	977	819 (17%)	USATODAY.COM	185	120 (36%)
INSTAGRAM.COM	261	169 (36%)	WARRIORFORUM.COM	346	312 (10%)
LEMONDE.FR	305	203 (34%)	WEATHER.COM	516	454 (13%)
LIBERO.IT	396	297 (25%)	WIX.COM	500	433 (14%)
LIFEHACKER.COM	86	66 (24%)	WORDPRESS.COM	63	33 (48%)
LINKEDIN.COM	185	146 (22%)	WORDREFERENCE.COM	185	120 (36%)
LIVE.COM	300	266 (12%)	XDA-DEVELOPERS.COM	689	566 (18%)
LIVEJOURNAL.COM	106	92 (14%)	YAHOO.COM	166	94 (44%)
MARCA.COM	723	515 (29%)	YOUTUBE.COM	325	296 (9%)
MASHABLE.COM	145	101 (31%)	ZIPPYSHARE.COM	563	470 (17%)
AVERAGE		22%			
AVERAGE ± STDEV		20%			
MEDIAN		20%			

Tabla 2
SCRIPTS EJECUTADOS EN PARALELO

	TOTAL SCRIPTS	PARALELO SCRIPTS	SAFE SCRIPTS
360.CN	27	12 (44%)	3 (11%)
ALIBABA.COM	12	7 (58%)	9 (75%)
ALLEGRO.PL	35	15 (42%)	9 (25%)
AMAZONWS.COM	14	9 (64%)	3 (21%)
BBC.COM	101	76 (75%)	15 (14%)
BET365.COM	22	12 (54%)	21 (95%)
BILD.DE	117	35 (29%)	16 (13%)
BLOGGER.COM	13	7 (53%)	6 (46%)
BLOGSPOT.COM	14	6 (42%)	6 (42%)
BLOOMBERG.COM	44	29 (65%)	32 (72%)
BOOKING.COM	36	5 (13%)	9 (25%)
CNET.COM	22	6 (27%)	7 (31%)
ENGADGET.COM	42	25 (59%)	19 (45%)
FORBES.COM	50	21 (42%)	17 (34%)
GITHUB.COM	4	3 (75%)	4 (100%)
GIZMODO.COM	18	8 (44%)	10 (55%)
GSMARENA.COM	77	11 (14%)	22 (28%)
IGN.COM	125	42 (33%)	33 (26%)
IKEA.COM	6	4 (66%)	5 (83%)
IMGUR.COM	41	17 (41%)	17 (41%)
INDIATIMES.COM	76	36 (47%)	13 (17%)
INSTAGRAM.COM	10	4 (40%)	4 (40%)
LEMONDE.FR	54	26 (48%)	4 (7%)
LIBERO.IT	40	16 (40%)	14 (35%)
LIFEHACKER.COM	18	8 (44%)	10 (55%)
LINKEDIN.COM	15	6 (40%)	6 (40%)
LIVE.COM	16	7 (43%)	9 (56%)
LIVEJOURNAL.COM	20	7 (35%)	10 (50%)
MARCA.COM	210	125 (59%)	136 (64%)
MASHABLE.COM	8	4 (50%)	5 (62%)
MEDIAFIRE.COM	28	15 (53%)	7 (25%)
PETFLOW.COM	32	22 (68%)	7 (21%)
PINTEREST.COM	8	2 (25%)	4 (50%)
REDIFF.COM	27	11 (40%)	12 (44%)
REUTERS.COM	95	15 (15%)	10 (10%)
RT.COM	34	17 (50%)	5 (14%)
SCRIPBD.COM	30	9 (30%)	9 (30%)
SOFTONIC.COM	27	11 (40%)	4 (14%)
SOURCEFORGE.NET	44	22 (50%)	16 (36%)
SPEEDTEST.NET	64	18 (28%)	10 (15%)
STACKEXCHANGE.COM	16	8 (50%)	2 (12%)
TAOBAO.COM	22	12 (54%)	15 (68%)
TARINGA.NET	62	37 (59%)	25 (40%)
TECHCRUNCH.COM	26	21 (80%)	16 (61%)
THEFREEDICTIONARY.COM	86	26 (30%)	26 (30%)
TIME.COM	33	19 (57%)	17 (51%)
TRIPADVISOR.COM	46	26 (56%)	44 (95%)
TUMBLR.COM	36	8 (22%)	11 (30%)
UPLOADED.NET	16	7 (43%)	2 (12%)
UPS.COM	95	73 (76%)	31 (32%)
USATODAY.COM	6	3 (50%)	6 (100%)
WARRIORFORUM.COM	27	8 (29%)	7 (25%)
WEATHER.COM	64	16 (25%)	11 (17%)
WIX.COM	23	7 (30%)	4 (17%)
WORDPRESS.COM	16	10 (62%)	14 (87%)
WORDREFERENCE.COM	8	5 (62%)	4 (50%)
XDA-DEVELOPERS.COM	46	19 (41%)	9 (19%)
YAHOO.COM	42	13 (30%)	17 (40%)
YOUTUBE.COM	23	10 (43%)	9 (39%)
ZIPPYSHARE.COM	27	15 (55%)	8 (29%)
AVERAGE		45%	40%
AVERAGE ± STDEV		45%	37%
MEDIAN		44%	35%

desviación típica es también del 45%. Mencionar que cuando un script está listo para ejecutarse, se le puede asignar al thread principal cuando se da la circunstancia de que este script es el primero de la cola y el thread principal queda

libre y solicita un nuevo evento. El número de scripts completamente seguros varía desde el 7% hasta el 100% en el mejor de los casos. La media es del 40%, la mediana es el 35% y la media descartando los valores que quedan fuera del rango de la media \pm la desviación típica es del 37%.

VI. TRABAJOS RELACIONADOS

La mayoría de los sistemas de automatización Web (*SmartBookmarks* [1], *Kapow* [2], *Wargo* [3], *Sahi* [4], *Selenium* [5], *WebMacros* [17], *WebVCR* [18]) utilizan el API de los navegadores convencionales para construir componentes de navegación capaces de ejecutar secuencias. Esta aproximación tiene dos ventajas importantes. Por un lado, no requiere del costoso desarrollo de un componente de navegación desde cero y por otro lado, al utilizar un navegador convencional, está garantizado que éste se comportará de la misma manera que cuando lo utiliza una persona. Por el contrario, esta aproximación puede no ser la más adecuada en entornos que requieren tiempos de respuesta muy reducidos o cuando se requiere de un gran volumen de navegadores ejecutándose al mismo tiempo. Es necesario incidir en que los navegadores convencionales son aplicaciones cliente y no han sido desarrollados pensando en tareas de automatización Web.

Otros sistemas utilizan la aproximación de desarrollar navegadores ligeros y muy simplificados. Este es el caso de *Jaunt* [8], sistema de automatización Web sin soporte para JavaScript.

Un último grupo de herramientas, proporciona una implementación completa que emula a algún navegador convencional (con soporte para JavaScript y otras funcionalidades avanzadas). Este es el caso de *EnvJS* [6] y *HtmlUnit* [7]. Estos sistemas suelen ser más eficientes que los navegadores convencionales dado que no están orientados a ser utilizados por personas y por lo tanto pueden evitar ciertas tareas que se ejecutan de manera incondicional en los navegadores convencionales (por ejemplo, la renderización del documento). Sin embargo, estos navegadores ad-hoc, operan de manera muy similar a cómo lo hacen los navegadores convencionales cuando construyen la representación interna de los documentos HTML, siendo éste un proceso secuencial.

Los navegadores convencionales evalúan los contenidos JavaScript de manera secuencial (salvo los dos casos excepcionales comentados en la sección 3.A). Aun así, estos navegadores ofrecen otras optimizaciones encaminadas a mejorar el rendimiento (ej. la precarga especulativa de *Mozilla Firefox* [13]). Existen otras herramientas de navegación que también explotan diferentes mecanismos de optimización. Por ejemplo *ZOOMM* [19] es un navegador paralelo que explota precarga de recursos, aplicación de estilos CSS en paralelo y compilación de scripts en paralelo (la evaluación se realiza de manera secuencial). *Adrenaline* [20], es un sistema que optimiza el proceso de carga dividiendo el documento en subdocumentos y procesando cada uno de estos fragmentos en un thread diferente.

VII. CONCLUSIONES

En este trabajo se presenta una novedosa técnica de optimización para un navegador especializado en la ejecución de secuencias de navegación Web. Esta técnica permite la

evaluación en paralelo de scripts contenidos en un mismo documento HTML, cuando estos scripts no tienen dependencias entre sí.

La aproximación utilizada se basa en recopilar cierta información relevante en una primera ejecución de la secuencia. En esta ejecución inicial, se monitoriza la evaluación del código JavaScript y se obtiene información sobre las interacciones entre los nodos que forman parte del árbol DOM. Con esos datos se construye un grafo de dependencias entre scripts que contiene, para cada script S contenido en el documento HTML, una lista con los otros scripts que deben evaluarse con anterioridad para que S se ejecute con éxito. Este grafo de dependencias entre scripts se guarda utilizando expresiones XPath para representar a los script que forman parte de él. En las siguientes ejecuciones de la misma secuencia se utiliza este grafo para determinar qué scripts se pueden evaluar en paralelo.

Los experimentos realizados sobre sitios Web reales, muestran cómo un alto porcentaje de los scripts se pueden evaluar en paralelo. Además, el componente de navegación construye los documentos HTML de manera más rápida cuando se aplican estos algoritmos.

REFERENCIAS

- [1] Hupp D., Miller R.C.: Smart Bookmarks: automatic retroactive macro recording on the web. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, pp. 81-90. ACM New York, Newport (2007).
- [2] Kapow, <http://kapowsoftware.com/>.
- [3] Pan A., Raposo J., Álvarez M., Hidalgo J., Viña A.: Semi automatic wrapper-generation for commercial web sources. In: IFIP WG8.1 Working Conference on Engineering Information Systems in the Internet Context, pp. 265-283. Kluwer, B.V. Deventer, Japan (2002).
- [4] Sahi. <http://sahi.co.in/w>.
- [5] Selenium, <http://seleniumhq.org>.
- [6] EnvJS. <http://www.envjs.com>.
- [7] HtmlUnit, <http://htmlunit.sourceforge.net>.
- [8] Jaunt Java Web Scraping & Automation. <http://jaunt-api.com>.
- [9] Losada J., Raposo J., Pan A., Montoto P.: Efficient execution of web navigation sequences. World Wide Web. DOI 10.1007/s11280-013-0259-8. ISSN 1386-145X.
- [10] Document Object Model (DOM). <http://www.w3.org/DOM>.
- [11] Grosskurth, A., Godfrey, M. W., September 2005. A reference architecture for web browsers. En: ICSM'05: Proceedings of the 21st IEEE International Conference on Software Maintenance (ICSM'05). pp. 661-664.
- [12] Simple API for XML. <http://sax.sourceforge.net>.
- [13] Mozilla HTML5 Parser. https://developer.mozilla.org/en-US/docs/Web/Guide/HTML/HTML5/HTML5_Parser.
- [14] HTML5. <https://html.spec.whatwg.org>.
- [15] XML Path Language (XPath), <http://www.w3.org/TR/xpath>.
- [16] Alexa, The Web Information Company. <http://www.alexa.com>.
- [17] Safonov A., Konstan J., Carlis J.: Beyond Hard-to-Reach Pages: Interactive, Parametric Web Macros. In: 7th Conference on Human Factors & the Web. Madison 2001.
- [18] Anupam V., Freire J., Kumar B., Lieuwen D., Automating web navigation with the WebVCR, Computer Networks 33(1-6), 503-517 (2000)
- [19] Calin Cascaval, Seth Fowler, Pablo Montesinos-Ortego, Wayne Piekarski, Mehrdad Reshadi, Behnam Robotmili, Michael Weber, and Vrajesh Bhavsar. 2013. ZOOMM: a parallel web browser engine for multicore mobile devices. En Proceedings of the 18th ACM SIGPLAN symposium on Principles and practice of parallel programming (PPoPP '13). ACM, New York, NY, USA, 271-280.
- [20] H. Mai, S. Tang, S. T. King, C. Cascaval, and P. Montesinos. A case for parallelizing web pages. In Proceedings of the 4th USENIX conference on Hot Topics in Parallelism, HotPar'12, Berkeley, CA, USA, June 2012. USENIX Association.

Monitorización y Análisis de Tráfico de Red con Apache Hadoop

Rubén García-Valcárcel, Rafael Leira, Iván González, Jorge E. López de Vergara
Departamento de Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior
Universidad Autónoma de Madrid
ruben.garcia-val@estudiante.uam.es, {rafael.leira, ivan.gonzalez, jorge.lopez_vergara}@uam.es

Resumen—La monitorización y el análisis del tráfico de red cobran gran importancia cuando se trata de optimizar los recursos de red y mejorar la experiencia de los usuarios. El uso de un único servidor que, a pesar de ser de alto rendimiento, consta de unos recursos limitados, no es escalable para el análisis exhaustivo de grandes volúmenes de datos. En este trabajo se presenta una herramienta basada en Apache Hadoop para la captura, el almacenamiento, el procesado y el análisis de grandes cantidades de tráfico de red. El ecosistema Hadoop proporciona una plataforma de computación escalable y fiable, capaz de integrar múltiples fuentes de datos con diferentes formatos. La herramienta desarrollada permite realizar análisis complejos sobre los datos de red de forma cómoda mediante un lenguaje basado en Structured Query Language (SQL), facilitando así la labor de los administradores de red. Los resultados muestran que, con un pequeño clúster de bajo coste, es posible realizar estas tareas de una forma eficiente, mejorando la capacidad y escalabilidad de las herramientas más potentes en la actualidad.

Palabras clave—Big Data, Hadoop, computación distribuida, PCAP, flujos de red, análisis de datos

I. INTRODUCCIÓN

Existen multitud de técnicas para la monitorización y el análisis del tráfico de red, pero fundamentalmente se pueden clasificar en activas o pasivas. La monitorización activa consiste en introducir tráfico en la propia red (por ejemplo, trenes de paquetes [1]) para realizar mediciones sobre el estado de la misma, lo que supone en algunos casos interferir con el tráfico legítimo que la atraviesa [2]. Además, este tipo de técnicas son incapaces de estudiar el comportamiento de los usuarios, así como de otros eventos no previstos en la red. Por otro lado, la monitorización pasiva consiste en capturar el tráfico y analizarlo para obtener información sin perturbar el comportamiento de la red. Es por ello que este tipo de monitorización suele preferirse a la hora de trabajar en la planificación y la gestión de redes de cierto tamaño. A raíz de esto, han surgido multitud de herramientas de monitorización pasiva. Algunos de los ejemplos más conocidos son: *Tcpdump* [3], *WireShark* [4], *CoralReef* [5], o *NetFlow* de Cisco [6].

Uno de los problemas de la monitorización pasiva es operar a la velocidad de la red, particularmente en los casos de redes de alta velocidad, ya que se son necesarios servidores de altas prestaciones para capturar y procesar los datos a la tasa del enlace. Si se desea almacenar los paquetes para un análisis más completo, se requiere de una gran capacidad de almacenamiento que opere a alta velocidad. Debido al crecimiento exponencial de las redes de comunicaciones, los requisitos de estos equipos son cada vez mayores, y del mismo modo, muchas de las herramientas mencionadas anteriormente no son capaces de trabajar con estas cantidades

de datos dada su limitada escalabilidad. Por ello, muchos de los nuevos desarrollos se están centrando en aprovechar todo el potencial de las plataformas hardware de alto rendimiento modernas, que permiten un procesamiento en paralelo de los datos [7]–[9]. Sin embargo, el uso de este tipo de hardware no resuelve todos los inconvenientes inherentes a las redes de alta velocidad, en especial la problemática del almacenamiento de los datos de red, que resulta de vital importancia. Disponer de toda la información posible permite realizar un análisis detallado del estado de la red, detectando así intrusiones, errores en la configuración, fallos en los componentes de red, ataques por denegación de servicio o incluso contingencias a nivel de aplicación o futuros problemas de escalabilidad.

Procesar grandes cantidades de datos (del orden de decenas de TeraBytes) requiere de una infraestructura adecuada, así como disponer de herramientas capaces de escalar al ritmo de crecimiento de los datos de red que cubran las necesidades de los administradores de red, en cuanto a simplicidad y funcionalidad. Así mismo, es importante la redundancia de los datos y la tolerancia frente a fallos del sistema sobre el que se ejecutan las nuevas herramientas. Si bien, actualmente existen nuevos paradigmas de computación y almacenamiento que cumplen con las características antes mencionadas, en este trabajo nos hemos centrado en el ecosistema de herramientas creado alrededor del proyecto Apache Hadoop. Para evaluar esta tecnología dentro del contexto de la monitorización de redes de comunicaciones, se ha puesto en marcha una maqueta de captura y procesamiento de tráfico de red dentro de la red local de laboratorios de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid.

El procesamiento de tráfico de red mediante las diversas herramientas del ecosistema Hadoop ha sido previamente tratado en la literatura [10], [11]. No obstante, dichas implementaciones sufren de diversas limitaciones, desde el rendimiento alcanzado y la baja escalabilidad, hasta limitaciones a la hora de procesar paquetes, y formatos. A lo largo de este trabajo se discuten dichas limitaciones, así como las diferentes mejoras que se han aplicado para solventarlas. Para ello, en la siguiente sección se analiza el estado del arte actual. En la sección III se detalla la arquitectura del sistema propuesto, junto con las diferentes tareas de preproceso y análisis realizadas. La sección IV muestra los resultados obtenidos, mientras que en la sección V se concluye este artículo y se presentan las líneas de trabajo futuro.

II. ESTADO DEL ARTE

El proyecto Apache Hadoop [12] nació como alternativa libre a las versiones MapReduce [13] y GFS [14] de Google.

El núcleo consta de un sistema de archivos distribuido, *Hadoop Distributed File System* (HDFS) [15] y del paradigma de programación *MapReduce*, conociéndose en su última versión estable como *YARN* o *MapReduce 2.0* [16]. Dichas herramientas permiten paralelizar el análisis de los datos de una forma transparente y eficiente, llevando la computación a los datos. El proceso es escalable y robusto, ya que si falla cualquier nodo, un disco o la red, el sistema continúa funcionando.

Sobre el núcleo de Hadoop surgen multitud de herramientas, comúnmente llamadas su ecosistema. Entre las herramientas más conocidas se encuentra Apache HBase para el almacenamiento estructurado de los datos y su posterior consulta, Apache Pig para el análisis a alto nivel, o Mahout para realizar aprendizaje automático. Dado que el tratamiento de los datos es un proceso complejo en sí mismo, es necesario proporcionar a los administradores de la red un medio que facilite este trabajo. Con esto en mente, se ha decidido utilizar la herramienta Apache Hive [17] que permite crear consultas en su propio lenguaje, HiveQL, basado en SQL (*Structured Query Language*), que se traducen automáticamente al paradigma MapReduce. Este ecosistema se está extendiendo día a día, llegando a grandes empresas como *eBay*, *Facebook*, *Google* o *LinkedIn* [18], llegando incluso a ofrecer una plataforma entera de computación sobre Hadoop. Algunos de los ejemplos más relevantes son la infraestructura *Elastic Compute Cloud* (EC2) de *Amazon*, *CDH* (*Cloudera's software distribution containing Apache Hadoop*) o el *IBM InfoSphere BigInsights* [19].

Hadoop fue diseñado para el procesamiento por lotes, con aplicaciones en diversos ámbitos como la minería de datos, la indexación de páginas web o el análisis de *logs*. Por ello, la lectura de archivos de texto plano está muy simplificada en Hadoop. Dentro del mundo de las redes y la monitorización, se considera al formato binario PCAP (*Packet CAPture*) [3] como un estándar *de facto* a la hora de almacenar paquetes capturados de la red. A pesar de que Hadoop también proporciona una clase que permite leer cómodamente archivos binarios, el formato PCAP no cumple con las restricciones impuestas por dicha clase. Por este motivo, los archivos PCAP no pueden ser utilizados de forma directa por Hadoop.

Una forma de resolver la cuestión anterior es realizar la conversión de formato para cada PCAP. No obstante, esto requiere demasiado tiempo y espacio, por lo que trabajos como los de RIPE [10] o Y. Lee [11] han creado sus propias clases Java que resuelven con este problema y permiten abrir archivos PCAP. El trabajo de RIPE ha sido uno de los primeros en mezclar el mundo de la computación distribuida y Hadoop con el mundo de las redes y su monitorización. Su implementación permite extraer la información de los paquetes, proporcionando una API con la información de los diversos niveles de la torre OSI. No obstante, esta implementación solo permite diseccionar los protocolos DNS y HTTP del nivel de aplicación, para lo cual utiliza bibliotecas de terceros. Dichas bibliotecas resultan muy pesadas por lo que la implementación de RIPE no alcanza un elevado número de Gbps al procesar tráfico.

En [11] Y. Lee presenta un trabajo un sistema de monitorización capaz de procesar, mediante una API diferente, tráfico IP, TCP, HTTP y registros NetFlow. Por último, en [20], Y. Lee presenta una heurística por la cual es posible fragmentar

archivos PCAP en diversos bloques, de modo que Hadoop pueda procesar un único archivo de varios GigaBytes o TeraBytes entre diferentes nodos. La implementación de Y. Lee reensambla los paquetes a nivel de flujo para procesarlos posteriormente. Este hecho impide su aplicación a nuestro caso, pues trabajamos sobre paquetes truncados.

Dentro de la computación distribuida existen otras aproximaciones orientadas al procesamiento de *streams* (flujos) en vez de por lotes. Por ejemplo, Apache Storm es una de las herramientas más conocidas de procesamiento en *streams* [21]. No obstante, un sistema basado en el procesado de *streams* plantería realizar el proceso deseado en el momento de la captura. Dado que el objetivo de este trabajo es la realización de análisis complejos bajo demanda, tales como estudios longitudinales del comportamiento de la red durante un año, el paradigma del procesamiento por lotes parece encajar mejor. En cualquier caso, en un escenario en el que hubiese que combinar procesos por lotes y *streams*, sería interesante el uso de una arquitectura lambda como la propuesta en [22].

III. DESARROLLO

En esta sección se describe la arquitectura del sistema propuesto. Se comienza detallando las características de la red que ha motivado este trabajo, ya que las mismas se han tenido en cuenta para la toma de algunas decisiones de diseño.

III-A. Caso de estudio

La solución propuesta en este trabajo se ha instalado y probado en la red de área local de los laboratorios docentes de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid. La topología de la red se muestra en la Fig. 1. La red consta de más de 700 ordenadores, que generan desde 150 hasta 260 GB de tráfico cada día lectivo, con un enlace de salida a Internet de 1 Gbps. Un cortafuegos bloquea los protocolos diferentes de DNS, HTTP y HTTPS para la mayor parte de los ordenadores, por lo que los análisis propuestos se centran en estos protocolos. Según se ilustra en la Fig. 1, para capturar el tráfico se ha dispuesto una sonda en la interfaz interna del cortafuegos. La arquitectura de captura podría ampliarse a varias sondas en un escenario en el que existiesen múltiples subredes independientes. Finalmente, la subred monitorizada cuenta con un servidor de DNS que gestiona las peticiones de la red interna. Al encontrarse este *proxy* dentro de la subred, la sonda de captura es incapaz de escuchar las peticiones DNS de los diferentes ordenadores, pudiendo tan solo capturar eventualmente las peticiones del propio servidor de DNS. Debido a esto no se conocen, a priori, todas las peticiones DNS realizadas, ni las direcciones de los ordenadores que las solicitan.

III-B. Arquitectura

La arquitectura del sistema propuesto realiza el procesamiento en tres fases bien diferenciadas, tal y como se puede observar en la Fig. 2. En la primera parte, una sonda captura los paquetes que circulan por la subred que se pretende monitorizar. Los paquetes son almacenados en diferentes archivos PCAP, mientras en paralelo se reconstruyen los flujos subyacentes. Posteriormente, estos datos son movidos a HDFS. En la segunda parte los archivos en formato PCAP son preprocesados mediante tareas *MapReduce* con el fin

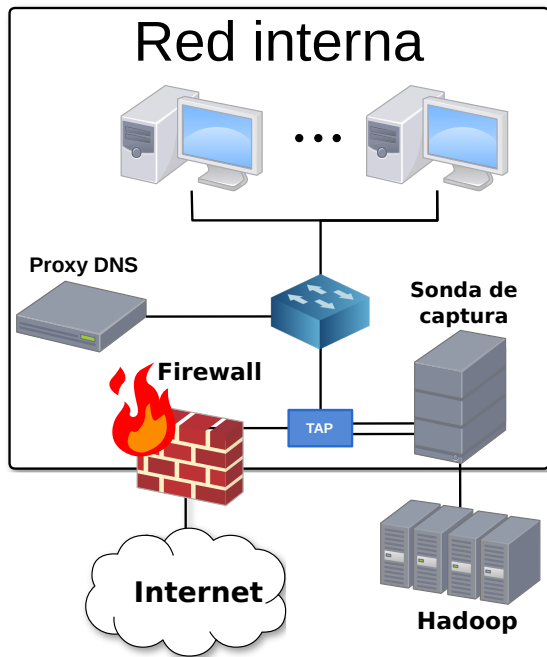


Figura 1. Subred analizada

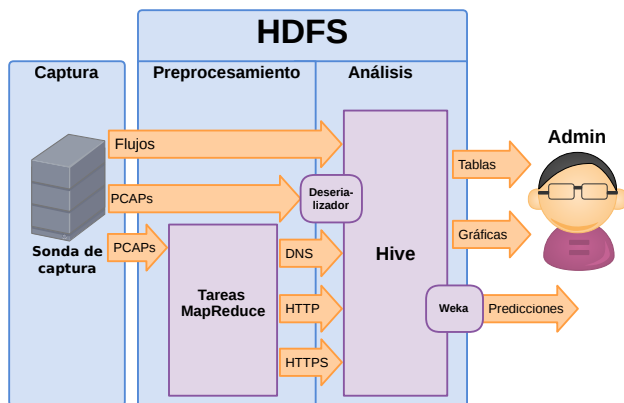


Figura 2. Arquitectura del sistema propuesto

de obtener información estructurada y poder realizar los posteriores análisis con mayor agilidad. Por último, en una tercera parte se realiza el análisis de los datos, para el cual se propone el uso Hive, que permite realizar consultas SQL sobre los datos ya procesados. En esta etapa el administrador de la red toma un papel fundamental, ya que puede tanto ejecutar alguna de las consultas que se proponen en este trabajo, como crear otras nuevas en función de sus necesidades. Las consultas más generales pueden hacerse sobre los archivos de flujos, pues contienen información resumida del tráfico de red. Un análisis más específico se puede llevar a cabo si se utilizan los archivos de los protocolos DNS, HTTP y HTTPS generados por las tareas *MapReduce*. Si aún así se necesita información más detallada, se puede utilizar un deserializador que transforma los paquetes en datos estructurados. De este modo, los archivos PCAP resultan transparentes para la otras herramientas convirtiéndose en una única tabla.

III-C. Captura

Para poder realizar un posterior análisis en profundidad, es necesario capturar y almacenar todos los paquetes que transcurren por la red. Para lograrlo, se ha recurrido a la aplicación FlowProcess [23], que permite guardar los paquetes capturados en archivos binarios con formato PCAP y al mismo tiempo, es capaz de reconstruir flujos unidireccionales y exportar su información en formato CSV. No obstante, se han realizado varias modificaciones a este programa para la adaptar los resultados al problema en cuestión. Algunas de las modificaciones más destacables son las siguientes:

- Almacenar toda la carga útil de los flujos supone un consumo de disco no despreciable, y necesita una velocidad de escritura elevada para que no suponga un cuello de botella. Para evitar este tipo de problemas, los paquetes mayores de 300 Bytes se truncan a esta cantidad. Se ha considerado que esta cifra mantiene un compromiso entre almacenamiento y capacidad de análisis, ya que en la mayoría de casos los 256 primeros bytes de la carga útil de un flujo son suficientes para clasificarlo mediante DPI (Inspección Profunda de Paquetes o Deep Packet Inspection) [24]. Por otro lado, este truncamiento asegura parcialmente la confidencialidad de los datos, a la vez que se consigue una reducción de algo más del 75 % en el espacio utilizado. Con la reducción propuesta, el sistema almacena actualmente 13 TB de archivos PCAP capturados a lo largo de un año. Dicho sistema, cuenta con la tecnología para capturar y almacenar hasta 10 Gbps sin pérdidas [25].
- La división de archivos PCAP no es una tarea sencilla ni liviana [20], ya que los paquetes que los forman son de longitud variable, y no hay ningún símbolo que separe los diferentes paquetes, como es el caso del carácter de salto de línea en los archivos de texto. Dado que tanto Hadoop como el paradigma *MapReduce* se fundamentan en el desplazamiento de la computación a los datos, es imprescindible disponer de la información repartida entre los múltiples nodos. Por tanto, realizar la división de los archivos en esta primera parte de captura es la solución más simple y con menor coste computacional. El tamaño recomendado de cada bloque está en torno a 100MB [26], por lo que el valor por defecto de este parámetro suele ser 64 o 128MB. Un tamaño mucho menor hace que los tiempos de búsqueda de archivos y de planificación de tareas se disparen. Por otro lado, un tamaño mucho mayor puede hacer que la asignación de tareas locales a los datos se convierta en una tarea costosa. Esto es debido al incremento de la probabilidad de que el conjunto de archivos sobre los que se va a trabajar no se encuentre distribuido de forma uniforme entre los diferentes nodos de cómputo. Si esto ocurre, algunos nodos pueden verse obligados a solicitar archivos a otros nodos reduciendo consecuentemente el rendimiento. Sin embargo, el tamaño recomendable está pensado para clústeres de decenas o cientos de nodos. En nuestro caso, al disponer de cinco nodos y varios TB de información a procesar, se ha observado que hasta un tamaño de bloque de 1 GB la localidad de los datos no supone un problema. Por ello, la sonda genera

archivos de este tamaño y no será necesaria una división en diferentes bloques de los mismos [20].

- La transferencia de los datos desde la sonda hasta el clúster de Hadoop genera tráfico en la red, que se ve forzado a pasar por la propia subred monitorizada y por tanto son capturados. Para evitar esta retroalimentación, se estudió establecer un enlace dedicado o una VLAN hacia el clúster de Hadoop, pero no fue posible por limitaciones en la infraestructura disponible. Por ello, se modificó la herramienta para que descartase todos aquellos paquetes que contengan la dirección IP de la sonda.

Finalmente, los archivos generados se copian al clúster de Hadoop durante la noche, aprovechando la reducción en la actividad de la red. Una vez se ha completado el proceso de captura, es posible pasar a la parte de preprocesado.

III-D. Preprocesado

La parte de preprocesado parte de la implementación de RIPE [10], sobre la cual se han realizado varios cambios. Estos son consecuencia de una adaptación a nuestro problema, además de un intento de mejora de rendimiento a costa de una pérdida de generalidad y flexibilidad. En primer lugar, la implementación de RIPE utiliza bibliotecas de terceros para procesar los protocolos DNS y HTTP que descartan todos los paquetes truncados, como es el caso de la mayor parte de los paquetes capturados. Por ello, se han desarrollado nuevos disectores desde cero de una forma más eficiente, guardando únicamente la información que se considera imprescindible.

Una ventaja de la implementación de RIPE es su flexibilidad y generalidad, la cual se obtiene gracias a la forma de guardar la información: en lugar de almacenar los campos en una estructura estática con registros predeterminados, la implementación de RIPE utiliza un *array* asociativo, cuyos índices vienen determinados por los nombres de los respectivos campos. Esto permite adaptar dinámicamente los campos de las estructuras a la información específica que contiene un determinado paquete o flujo. Un ejemplo práctico es la disección de HTTP: la cabecera tiene una gran variedad de campos posibles que además evolucionan con el tiempo [27]. En nuestro caso de estudio conocemos a priori todos los campos que nos interesan, por lo que una implementación como esta supone una pérdida de eficiencia tanto de memoria como de tiempo. Por este motivo nuestra implementación almacena la información en clases solo con los campos apropiados.

Otra ventaja de su implementación es la posibilidad de añadir cómodamente nuevas clases que amplíen el repertorio de protocolos que pueden ser diseccionados sin modificar su código. Esta flexibilidad se consigue indicando al principio, mediante un parámetro de Hadoop, el nombre de la clase que se va a utilizar para procesar la capa de aplicación. El principal inconveniente de esta aproximación es que solo se puede especificar un protocolo, por lo que si se quieren analizar varios se tendrá que codificar un programa *MapReduce* por cada protocolo. En nuestro caso se ha mantenido dicho diseño y, aprovechando la flexibilidad que proporciona se ha añadido una clase que disecciona HTTPS, dado que una gran parte del tráfico web de hoy en día se encuentra protegido por este protocolo. Sin embargo, con vistas a mejorar la eficiencia, en

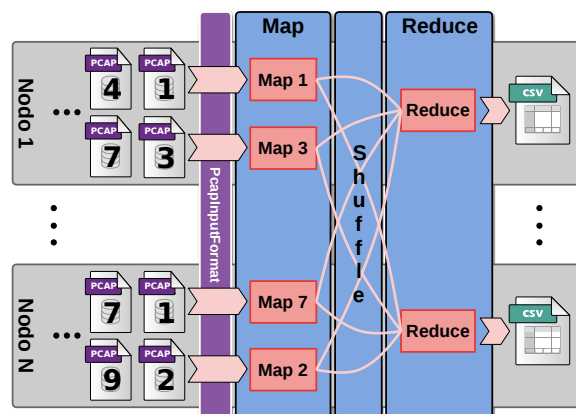


Figura 3. Ejemplo de ejecución de un programa *MapReduce* sobre archivos PCAP.

un futuro se planea crear una única clase capaz de analizar todos los protocolos a nivel de aplicación simultáneamente.

Otro cambio importante sobre el diseño de RIPE se ha hecho en el retorno de los paquetes: mientras que su iterador devuelve todos los paquetes que lee, el nuestro no devuelve aquellos que no contienen información sobre el protocolo indicado al inicio. Como esto se va a utilizar para el análisis de los protocolos de nivel de aplicación, vemos innecesaria la obtención de paquetes que no contienen esta información.

Por último, en la implementación original de RIPE, se aprecia una gran pérdida de rendimiento ocasionada por el reensamblado de los paquetes a nivel IP y la reconstrucción de flujos TCP unidireccionales. Debido a que nuestro tráfico se encuentra truncado y solo nos interesan los primeros bytes de cada flujo, en la mayoría de los casos podremos prescindir del reensamblado.

III-D1. Disectores: Una vez se han enumerado las diferencias con la aproximación inicial de RIPE, es posible comenzar a explicar los detalles de la implementación de los diferentes disectores de tráfico. En primer lugar, se dará una visión general de su funcionamiento, que en gran medida es la forma de actuar del paradigma *MapReduce*. En la Fig. 3 se muestra un diagrama simplificado de las fases de la ejecución de los disectores realizados. El planificador de Hadoop asigna qué archivo debe procesar cada nodo, de forma que se lance una tarea *Map* por archivo. Gracias a la clase `PcapInputFormat`, las tareas *Map* reciben solo los paquetes del protocolo especificado, que procesan por separado y emiten uno o varios pares (clave, valor) por cada paquete. En la fase de *Shuffle*, interna de Hadoop, los pares con la misma clave son agrupados y enviados al nodo adecuado para poder iniciar la última etapa. En esta última fase de *Reduce*, se procesan por separado cada uno de los grupos de valores que comparten la clave, generándose una o varias líneas de texto de la salida.

En las siguientes subsecciones, se comentan en detalle cada uno de los disectores realizados, así como la salida que estos generan.

III-D2. Disector HTTP: El disector HTTP empareja las peticiones HTTP con sus respectivas respuestas. La salida de

este programa es un archivo CSV con la información de cada flujo HTTP. Los campos contenidos en este archivo son: las direcciones IP, los puertos, los tiempos de inicio de la petición y de llegada de la respuesta, el método HTTP usado, el host de destino, el recurso de la URL solicitada, y por último el código y la descripción de la respuesta. Retransmisiones o peticiones sin respuesta son descartadas.

De la obtención de los campos HTTP se encarga la clase `HttpPcapReader`, implementada sobre la base de RIPE. Esta recibe la carga útil del protocolo de nivel de transporte, e identifica el protocolo de aplicación utilizando DPI y expresiones regulares simples. Al encontrarse los paquetes truncados, es posible que ciertas partes de la cabecera no estén presentes y otras se encuentren incompletas. Gracias a esta clase, las tareas *Map* procesan cada paquete identificado como HTTP de uno en uno. Para emparejar cada petición con su respuesta, estas tareas generan claves a partir de las direcciones IP y los puertos. Es importante recalcar que, aunque el número de *ACK* de la respuesta se puede calcular a partir de la petición, si la petición está segmentada a nivel TCP no se podrá realizar este cálculo, ya que no se realiza el reensamblado. En los valores se guardará el resto de campos necesarios para generar la salida anterior, incluyendo el *ACK* en caso de las respuestas y el siguiente número de secuencia en el de las peticiones. Finalmente, las tareas *Reduce* del disector ordenan los valores antes mencionados, generando una línea del archivo de salida por cada conexión HTTP, incluyendo por tanto, la información relevante de su petición y su respuesta.

III-D3. Disector DNS: Muchas de las peticiones HTTP capturadas no contienen el campo *Host* debido al truncado de los paquetes en la fase de captura. Consultar el nombre de todos los servidores mediante un script que acceda al DNS de la UAM es poco eficiente, y además dado el número de consultas por segundo a resolver, podría ser considerado un DDoS, causando un posible bloqueo temporal por parte del servidor. La solución propuesta es una búsqueda en el tráfico DNS disponible para identificar el nombre de los servidores a partir de su dirección IP. El objetivo de este disector es crear una tabla con todos los nombres consultados hasta el momento, junto con su resolución IP de forma similar a [28].

Sobre la base de RIPE se ha implementado la clase `DnsPcapReader`, que obtiene todas las respuestas a las consultas de tipo *A* (dirección de un host) y de clase *IN* (Internet). Aunque el registro *CNAME* es también importante para la resolución final, actualmente se está trabajando para soportar este tipo de respuestas y así ampliar el número de direcciones IP almacenadas. Todos los demás tipos de respuestas son ignoradas, así como todas las consultas, puesto que lo único que nos interesa es la resolución de las direcciones IP. Usando la clase `DnsPcapReader`, las tareas *Map* reciben las consultas DNS mencionadas y emiten, para cada una de sus respuestas, una cadena de texto que consta únicamente del nombre consultado y la dirección IP resuelta. Las tareas *Reduce* se encargan de eliminar las cadenas duplicadas.

III-D4. Disector HTTPS: Cómo resultado del aumento de la seguridad en los sistemas actuales, la mayor parte del tráfico se encuentra cifrado. El protocolo por excelencia en este campo es HTTPS, por lo que se ha considerado procesar también este protocolo. Aunque los datos se encuentran cifrados, el nombre del servidor se puede obtener fácilmente

examinando la negociación del protocolo TLS. Con este fin se ha implementado la clase `HttpsPcapReader` sobre la base de RIPE, que considera únicamente los mensajes de tipo *Hello*, y obtiene de ellos el nombre del servidor. Tras esto, la tarea *MapReduce* genera una línea por paquete procesado con las direcciones IP de las máquinas que intervienen, y el nombre del servidor obtenido.

IV. RESULTADOS

En esta sección se detallan las características del hardware utilizado para realizar las pruebas de rendimiento y se muestran los resultados de las mismas. Por último se especifican los análisis que se han realizado sobre los datos de los que se disponen, junto a algunas de las gráficas más interesantes que se han obtenido.

IV-A. Pruebas de Rendimiento

Antes de comparar los resultados del rendimiento de la herramienta presentada, en la Tabla I se detallan las características de los sistemas que se han utilizado para evaluarlo. Como se puede observar, se dispone de dos sistemas: un pequeño clúster con Apache Hadoop, y un ordenador de altas prestaciones donde se han realizado pruebas en serie (a partir de ahora, nodo serie).

El clúster de Hadoop consta de 8 nodos conectados mediante una red de 1 Gbps, de los cuales solo los 5 nodos más potentes realizan la computación. Por otro lado, para evaluar nuestros resultados se ha tenido en cuenta el rendimiento obtenido por Y. Lee [11], que dispone de un clúster de 30 nodos conectados mediante una red de 1 Gbps, cada uno con 8 cores a 2,93 GHz.

Para evaluar el rendimiento de los tres dissectores explicados, se ha comparado su velocidad de procesamiento con 1) un test de lectura incluido en la distribución de Hadoop, 2) una versión en serie escrita en C del disector HTTP [29] y ejecutada en el nodo serie, y 3) los resultados presentados por Y. Lee. La entrada de todas las pruebas realizadas ha sido la misma, 1 TB de tráfico real. Nótese que la comparación con las pruebas de Y. Lee no es totalmente justa, pues ni el tráfico ni el procesamiento realizado son los mismos. No obstante, sus diferentes análisis alcanzan rendimientos parecidos, y para la comparativa se han usado sus mejores tiempos.

En la Tabla II se muestra el rendimiento de todas estas pruebas, medido en Gbps y en Mpps (Millones de paquetes por segundo), así como su eficiencia medida en Gbps por core utilizado. Se puede observar que el rendimiento de cada core del sistema implementado dobla el de los cores de Y. Lee, quienes a su vez superan el rendimiento de RIPE [11].

Por otro lado, el test de lectura de Hadoop, que consiste en ejecutar el benchmark `TestDFSIO` disponible con Hadoop, tiene un rendimiento mucho menor que el de los dissectores, que deben procesar los datos además de leerlos. Esto se debe a que el test de lectura permite la división de los archivos en 8 bloques de 128 MB, por lo que el planificador de tareas necesita mucho más tiempo para completar su trabajo, convirtiéndose en el cuello de botella. Reduciendo el tamaño de la entrada se ha comprobado que los tiempos de lectura y de los dissectores se igualan, por lo que la mayor parte del tiempo se invierte en la lectura de los archivos. Finalmente, el principal hándicap del sistema propuesto es su heterogeneidad, puesto

Tabla I
CARACTERÍSTICAS DE LOS SISTEMAS DE PRUEBAS

Sistema	RAM (GB)	Disco (TB)	CPU	Total cores
Esclavo 1	32	15,2	1x Intel Xeon CPU L5408 @ 2,13GHz	4
Esclavo 2	32	15,2	1x Intel Xeon CPU L5408 @ 2,13GHz	4
Hadoop Esclavo 3	256	12,1	4x Intel Xeon CPU E7-4830 @ 2,13GHz	32
Esclavo 4	64	20,6	2x Intel Xeon CPU E5-2620 v3 @ 2,40GHz	12
Esclavo 5	64	20,6	2x Intel Xeon CPU E5-2620 v3 @ 2,40GHz	12
Serie	32	RAID0: 9x 3TB	2x Intel Xeon E5-2630 @ 2,6GHz	12

que el número de cores por máquina varía desde 4 hasta 32. Debido a esto, algunos nodos terminan sus tareas antes que otros, viéndose forzados a pedir a otros nodos nuevos archivos PCAP que procesar, lo que acaba por provocar una penalización por el aumento en las comunicaciones.

Tabla II
RENDIMIENTO DE CADA SISTEMA Y DE CADA CORE

Programa	Gbps	Mpps	Gbps/core
Hadoop Disector HTTP	7,13	3,96	0,11
Disector HTTPS	7,41	4,12	0,12
Disector DNS	7,72	4,29	0,12
Test de lectura	4,98	-	0,08
Disector HTTP en serie	4,21	2,34	4,21
Versión de Y. Lee [11]	5 nodos	1,9	0,05
	30 nodos	14,0	0,06

IV-B. Análisis

En esta sección se pretende mostrar algunos de los resultados que pueden ser generados por el conjunto de herramientas propuestas.

IV-B1. Consultas: Mediante HiveQL, el lenguaje de Hive basado en SQL, un administrador de red puede crear sus propias consultas que, de forma automática y transparente, son traducidas a tareas MapReduce. Se dispone así de una forma cómoda de realizar un análisis de grandes cantidades de datos. En la Fig. 4 se muestra un ejemplo de consulta que obtiene las peticiones HTTP más solicitadas, cuyos resultados han servido para generar la Fig. 6a.

```

1 select host, count(*) as numero
2 from http
3 group by host
4 order by numero desc;

```

Figura 4. Ejemplo de consulta en HiveQL

Sobre los archivos de flujos pueden realizarse las consultas más generales, es decir, las que requieran únicamente campos como direcciones IP, puertos, número de paquetes o duración del flujo. Por ejemplo, se han creado series temporales que muestran en cada momento el grado de utilización de la red, de una subred determinada, o incluso de cada uno de los ordenadores. Gracias a este tipo de consultas se han podido detectar ordenadores averiados, que no generan ningún tipo de tráfico a lo largo de un periodo largo de tiempo, así como los ordenadores que se han dejado encendidos alguna noche. Otros ejemplos relevantes son el estudio de los puertos y protocolos usados, ya que con esto se puede comprobar si el cortafuegos está funcionando correctamente, así como analizar las conexiones iniciadas desde el exterior con el fin de detectar

posibles ataques. En la Fig. 5 se muestra el uso de cada puerto y protocolo. El anillo exterior representa la cantidad de bytes, mientras que el anillo interno, muestra el número de flujos.

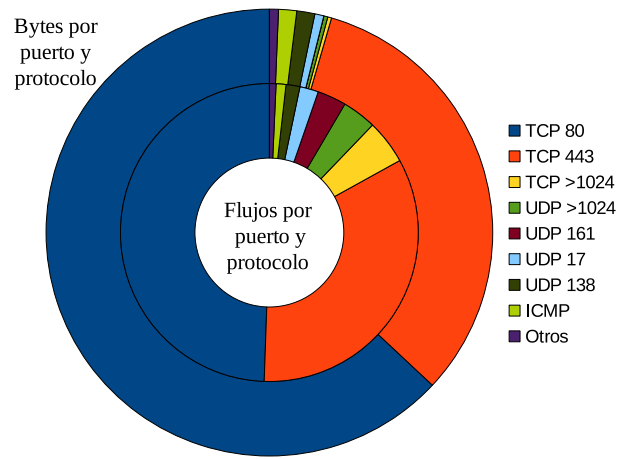


Figura 5. Porcentaje de Bytes (anillo exterior) y de flujos (anillo interior) de los puertos y protocolos más comunes

En el caso particular presentado, al tratarse de una red universitaria, se pueden tratar también de optimizar los recursos de red. Para ello se ha estudiado el consumo de ancho de banda de cada laboratorio en función de las horas del día, si hay o no clase y la asignatura que se imparte en cada momento. Es posible también estimar el número de ordenadores activos simultáneamente en días especiales como huelgas, periodos de exámenes o vacaciones con el fin de minimizar el personal o el número de laboratorios abiertos necesarios sin perjudicar a los usuarios.

Utilizando los archivos resultantes del preprocesamiento se puede realizar un análisis más específico, que tenga en cuenta campos de los protocolos DNS, HTTP o HTTPS. Usando los datos DNS se puede rellenar el campo Host de los registros HTTP truncados en los que se perdía esta información. Con los datos HTTP y HTTPS se han realizado diferentes tipos de consultas y análisis. Gracias a las consultas más simples, es posible obtener las URLs más populares, así como las páginas más visitadas durante cada asignatura. El top de URLs suele ser una métrica común a la hora de realizar diversos análisis red, ya que aporta interesante información acerca de utilización de la misma por parte de los usuarios. En nuestras pruebas, es observable que los usuarios dedican gran parte del tiempo en los laboratorios al ocio (ver Fig. 6). No obstante, a pesar de las ventajas que ofrece este tipo de análisis, realizarlo supone un proceso computacionalmente costoso, por lo que suele ser uno de los elementos más comparados. Por otro lado,

también se ha realizado un análisis más sofisticado con el objetivo de caracterizar el comportamiento de los usuarios en horario de clase. Este se explicará en la siguiente subsección.

Por último, si durante el análisis fuesen necesarios campos no disponibles en alguno de los archivos preprocesados, sigue siendo posible realizar las consultas directamente sobre los archivos PCAP mediante un deserializador.

IV-C. Predicciones

Un paso más allá en el análisis de la información disponible es la aplicación de técnicas de aprendizaje automático (*Machine Learning*) para obtener perfiles de uso de la red e incluso predicciones de comportamiento. Actualmente se está trabajando en caracterizar el comportamiento de los usuarios en horario de clase. Para ello en primer lugar es necesario analizar el tiempo transcurrido entre peticiones HTTP consecutivas para ser capaces de distinguir entre las peticiones voluntarias realizadas por los usuarios, de las secundarias realizadas por el navegador como consecuencia de las anteriores. Entendemos cómo petición secundaria, a aquellas peticiones que incluye el documento HTML de la petición inicial, cómo pueden ser imágenes, archivos Javascript o publicidad. Se ha comprobado que más del 90 % de las peticiones secundarias se hallan en los primeros 1,5 segundos, lo que hace posible un gran filtrado que reflejará mejor el comportamiento de los usuarios. Por otro lado, es necesario identificar cuándo los usuarios inician o cierran la sesión de trabajo, con el objetivo de diferenciar usuarios. Para ello, se ha usado una heurística que estima el momento en el que los usuarios inician o cierran la sesión, así como el sistema operativo que están usando. Esta se basa en los largos periodos de inactividad de los ordenadores, junto al hecho de que al encenderse, las máquinas solicitan determinadas URLs. Dichas URL dependen del sistema operativo que está iniciándose.

Un primer procesamiento mediante la herramienta de aprendizaje automático Weka [30] de los resultados obtenidos a partir de un conjunto reducido de datos, ha permitido confirmar que es posible caracterizar el comportamiento medio de los usuarios cuando hay clase. Muchas de las sesiones comparten un inicio en un intervalo corto de tiempo, así como una duración parecida. De forma similar, la gran mayoría comparten el sistema operativo. El siguiente paso es aplicar las mismas ideas sobre todos los datos disponibles, y en este caso, se empleará Apache Mahout [31].

V. CONCLUSIONES Y TRABAJO FUTURO

A lo largo de este trabajo se ha presentado una herramienta capaz de procesar grandes cantidades de datos obtenidos de la monitorización de redes de comunicaciones, permitiendo realizar diferentes tipos de análisis sobre la enorme cantidad de tráfico que circula por ellas, de una forma cómoda, eficiente y distribuida. Gracias al proyecto Apache Hadoop y al ecosistema que gira en torno a este, la herramienta desarrollada resulta escalable, tolerante frente a fallos, y facilita una paralelización del análisis transparente al gestor de la red. Con esta herramienta es posible cubrir el proceso de captura del tráfico de red, el preprocesado intermedio de los datos y un último y simplificado análisis mediante la herramienta Hive. El preprocesado se ha mostrado más eficiente que los trabajos previos que usan este tipo de arquitectura, alcanzando

una velocidad de procesamiento superior a 7 Gbps con un pequeño clúster. Por otro lado, el análisis propuesto ha permitido obtener datos interesantes sobre la red analizada, siendo además, fácilmente ampliable mediante nuevas consultas en un lenguaje basado en SQL.

A la monitorización de redes y el procesamiento distribuido de datos mediante Hadoop le queda un largo recorrido y desarrollo para llegar a alcanzar toda la funcionalidad y versatilidad del resto de herramientas actuales. Es por ello, que aun quedan múltiples opciones por explorar. Si bien una aproximación basada únicamente en procesamiento por lotes puede encontrarse limitada en rendimiento, es posible que una aproximación mixta entre un proceso en *streaming* de captura junto con un proceso bajo demanda por lotes, mejore tanto la velocidad como la escalabilidad del sistema final. Del mismo modo, se planea crear una única clase que unifique la disección de los diferentes protocolos de nivel de aplicación, ya que por el momento, solo se pueden obtener los registros de uno de los protocolos por cada ejecución. Dicha clase, se pretende ser reimplementada en C y unificada con el resto del proyecto mediante JNI, con el objetivo de mejorar el rendimiento de la aplicación. Además, se pretende profundizar más en la parte del análisis, mejorando las técnicas usadas de aprendizaje automático, así como integrando dichas técnicas con Hadoop a través de la herramienta Apache Mahout. Como último objetivo, se pretende enmarcar todo el conjunto de herramientas y clases desarrolladas en un marco para ofrecer a un administrador de red la capacidad de monitorizar y analizar una red, aunque dicha red supere los 10 Gbps, o incluso los 40 Gbps.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad de España (proyecto *PackTrack*, TEC2012-33754) y por la Universidad Autónoma de Madrid (proyecto “*Implementación de Modelos Computacionales Masivamente Paralelos*” (CEMU-2013-14)).

También queremos agradecer la colaboración de la Subdirección de Infraestructuras y de los técnicos de los laboratorios de Escuela Politécnica Superior de la Universidad Autónoma de Madrid.

REFERENCIAS

- [1] J. Ramos, P. M. Santiago del Río, J. Aracil, and J. E. López de Vergara, “On the effect of concurrent applications in bandwidth measurement speedometers,” *Computer Networks*, vol. 55, no. 6, pp. 1435–1453, 2011.
- [2] H. Veiga, T. Pinho, J. Oliveira, R. Valadas, P. Salvador, and A. Nogueira, “Active Traffic Monitoring for Heterogeneous Environments,” in *Networking - ICN 2005*, ser. Lecture Notes in Computer Science, P. Lorenz and P. Dini, Eds. Springer Berlin Heidelberg, 2005, vol. 3420, pp. 603–610.
- [3] “Tcpdump & libpcap.” [Online]. Available: <http://www.tcpdump.org/>
- [4] “Wireshark.” [Online]. Available: <https://www.wireshark.org/>
- [5] “Coralreef software suite.” [Online]. Available: <http://www.caida.org/tools/measurement/coralreef/>
- [6] “Cisco netflow.” [Online]. Available: <http://www.cisco.com/web/go/netflow>
- [7] R. Leira, P. Gómez, I. González, and J. E. López de Vergara, “Multi-media flow classification at 10 Gbps using acceleration techniques on commodity hardware,” in *Smart Communications in Network Technologies (SaCoNeT), 2013 International Conference on*, vol. 03, June 2013, pp. 1–5.

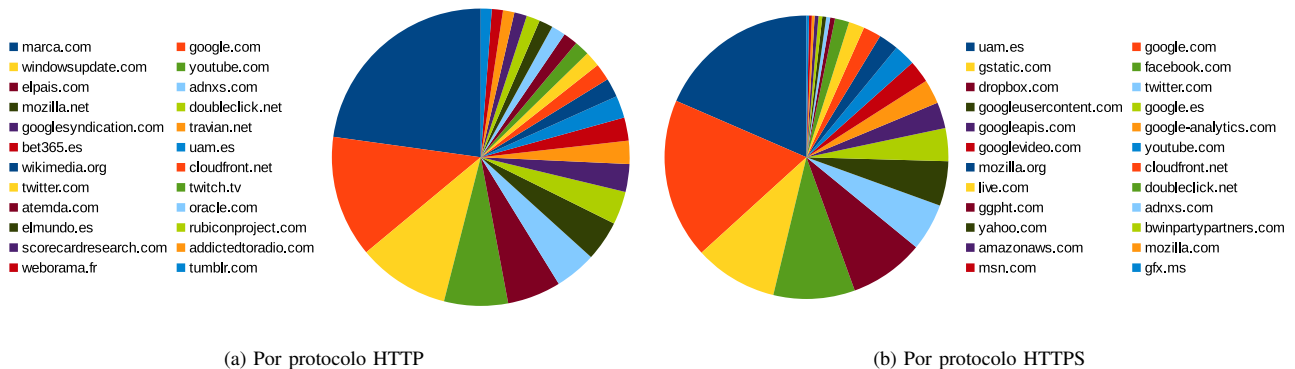


Figura 6. Top de peticiones a host por protocolo

[8] G. Vasiliadis, S. Antonatos, M. Polychronakis, E. Markatos, and S. Ioannidis, "Gnort: High Performance Network Intrusion Detection Using Graphics Processors," in *Recent Advances in Intrusion Detection*, ser. Lecture Notes in Computer Science, R. Lippmann, E. Kirda, and A. Trachtenberg, Eds. Springer Berlin Heidelberg, 2008, vol. 5230, pp. 116–134.

[9] J. J. Garnica, S. López-Buedo, V. López, J. Aracil, and J. M. Gomez Hidalgo, "A FPGA-based scalable architecture for URL legal filtering in 100GbE networks," in *ReConFig*, 2012, pp. 1–6.

[10] "Large-scale PCAP Data Analysis Using Apache Hadoop." [Online]. Available: <https://labs.ripe.net/Members/wnagele/large-scale-pcap-data-analysis-using-apache-hadoop>

[11] Y. Lee and Y. Lee, "Toward Scalable Internet Traffic Measurement and Analysis with Hadoop," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 1, pp. 5–13, January 2013.

[12] Apache Foundation, "Apache Hadoop." [Online]. Available: <http://hadoop.apache.org>

[13] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, January 2008.

[14] S. Ghemawat, H. Gobioff, and S. Leung, "The Google File System," *ACM SIGOPS Operating Systems Review*, vol. 37, no. 5, pp. 29–43, December 2003.

[15] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on*, May 2010, pp. 1–10.

[16] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth *et al.*, "Apache hadoop YARN: Yet Another Resource Negotiator," in *Proceedings of the 4th annual Symposium on Cloud Computing*. ACM, 2013, p. 5.

[17] A. Thusoo, J. S. Sarma, N. Jain, Z. Shao, P. Chakka, S. Anthony, H. Liu, P. Wyckoff, and R. Murthy, "Hive: a warehousing solution over a map-reduce framework," *Proceedings of the VLDB Endowment*, vol. 2, no. 2, pp. 1626–1629, August 2009.

[18] "Institutions using Hadoop." [Online]. Available: <https://wiki.apache.org/hadoop/PoweredBy>

[19] "Products that include Hadoop." [Online]. Available: <http://wiki.apache.org/hadoop/Distributions%20and%20Commercial%20Support>

[20] Y. Lee, W. Kang, and Y. Lee, "A Hadoop-based Packet Trace Processing Tool," in *Proceedings of the Third International Conference on Traffic Monitoring and Analysis*, ser. TMA'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 51–63. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1986282.1986289>

[21] P. Basanta-Val, N. Fernández-García, A. Wellings, and N. Audsley, "Improving the predictability of distributed stream processors," *Future Generation Computer Systems*, vol. 52, pp. 22 – 36, 2015, special Section: Cloud Computing: Security, Privacy and Practice. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167739X15000850>

[22] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co., 2015.

[23] P. M. Santiago del Río, "Internet traffic classification for high-performance and off-the-shelf systems," Ph.D. dissertation, Universidad Autónoma de Madrid, 2013.

[24] N. Cascarano, L. Ciminiera, and F. Risso, "Optimizing Deep Packet Inspection for High-Speed Traffic Analysis," *J. Netw. Syst. Manage.*, vol. 19, no. 1, pp. 7–31, Mar. 2011.

[25] V. Moreno, P. M. Santiago del Río, J. Ramos, D. Muelas, J. L. García-Dorado, F. J. Gómez-Arribas, and J. Aracil, "Multi-granular, multi-purpose and multi-Gb/s monitoring on off-the-shelf systems," *International Journal of Network Management*, vol. 24, no. 4, pp. 221–234, 2014.

[26] T. White, *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 2012.

[27] "Message Headers." [Online]. Available: <http://www.iana.org/assignments/message-headers/message-headers.xml>

[28] I. N. Bermudez, M. Mellia, M. M. Munafò, R. Keralapura, and A. Nucci, "Dns to the rescue: discerning content and services in a tangled web," in *Proceedings of the 2012 ACM conference on Internet measurement conference*. ACM, 2012, pp. 413–426.

[29] C. G. Vega Moreno, "Diseción de tráfico web a alta velocidad," Master's thesis, Universidad Autónoma de Madrid, Sep. 2014.

[30] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[31] Apache Foundation, "Apache Mahout." [Online]. Available: <http://mahout.apache.org>

Uso de información de la posición de los nodos en el protocolo de encaminamiento DYMO

Enrica Zola, Francisco Barceló-Arroyo, Israel Martín-Escalona

Departamento de Ingeniería Telemática,
Universidad Politécnica de Catalunya
c/ Jordi Girona 1-3, Barcelona 08034.
{enrica,barcelo,imartin}@entel.upc.edu

Abstract—Es un hecho conocido que el conocimiento de la posición de los nodos de una red inalámbrica es en general una herramienta útil para mejorar el funcionamiento de la misma. En el caso de las MANET, donde los algoritmos de encaminamiento presentan una mayor complejidad debido a la mayor variabilidad en la topología de la red, la mejora puede resultar muy notable. Protocolos como el Beaconless Routing (BLR) utilizan el conocimiento de las posiciones de los nodos para configurar los caminos de una manera distribuida. Este trabajo se apoya en ideas de BLR para presentar una nueva propuesta de DYMO. Los nodos calculan el retardo basándose en las posiciones y el nodo con menor retardo calculado es el que envía el Route Request (RREQ) en primer lugar. El resto de nodos descartan el envío del RREQ si lo han escuchado previamente. De esta forma, la selección del mejor nodo para encaminar se realiza de forma distribuida, esperándose por tanto una reducción significativa en la cantidad de RREQ transmitidos.

Index Terms—MANET, algoritmos de encaminamiento, DYMO, localización.

I. INTRODUCCIÓN

El encaminamiento es un punto importante de las redes MANET. Los cambios frecuentes en la topología de dichas redes complican el uso de algoritmos de encaminamiento pensados para redes más estables. En general las modificaciones planteadas para MANET presentan problemas de escalabilidad debido a que el aumento de los mensajes de control que deben transmitirse aumenta proporcionalmente con el número de combinaciones de rutas y no de manera lineal con el número de nodos. Una solución para MANET es el Dynamic MANET On-demand routing protocol (DYMO) [1]. Como protocolo reactivo, DYMO establece una ruta desde la fuente al destino antes de empezar la transmisión de datos. Una diferencia importante entre DYMO y aproximaciones más clásicas como AODV es que los nodos intermedios memorizan también la ruta a todos los nodos anteriores en el mismo camino, lo cual reduce el número de Route Request (RREQ) necesarios.

En la actualidad es frecuente encontrar dispositivos móviles con capacidades de localización. La información que proporcionan dichas capacidades puede ser utilizada para mejorar la eficiencia del encaminamiento. Diversas propuestas han sido presentadas con el objetivo de reducir el espacio de búsqueda en la fase de descubrimiento de rutas [2] o de aplicar el encaminamiento desde la fuente y así trazar el camino geográfico que los paquetes deben seguir [3]. En el Beaconless routing (BLR) [4] el nodo origen difunde paquetes de datos sin necesidad de conocer el camino de antemano. Por otra parte el BLR intercambia información de posición de

los nodos vecinos de manera periódica para mantener las tablas de encaminamiento actualizadas. Dicha periodicidad está condicionada por la precisión de la información de posicionamiento. De esta forma, cuanto mayor sea la precisión mayor será la frecuencia y por consiguiente el overhead. Debido a esto, los autores en [5] han propuesto el algoritmo Dynamic Route Maintenance para auto configurar el intervalo de envío del Beacon en función del patrón de movilidad de los nodos vecinos (ej. intervalos más cortos en caso de mayor movilidad).

El objetivo de este trabajo es valorar los beneficios de aplicar la estrategia del BLR al proceso de descubrimiento de rutas de DYMO. En el presente artículo se denomina a dicha propuesta DYMOselfwd (DYMO con forwarding selectivo) y se fundamenta en dos hipótesis: 1) los nodos conocen su propia posición en todo momento y 2) los nodos siempre conocen la posición del nodo destino. Al igual que en otros trabajos [4], [6], se asume la existencia de procedimientos de red que se encargan de proporcionar dichas posiciones. La incorporación de la información de posición en DYMOselfwd se espera que redunde en una reducción del overhead del protocolo. Específicamente 1) el número de nodos intentando acceder al medio para reenviar los RREQs será menor (es decir, menor probabilidad de colisión) y 2) se reducirá el tráfico en la red (es decir, menor congestión en la red). Se prevé además que la reducción en el número de colisiones compensará el incremento en el retardo asociado al DYMOselfwd, de forma que el retardo de establecimiento de ruta sea similar al presente en DYMO.

El uso de la información de localización en la etapa de descubrimiento de rutas de AODV se ha propuesto con anterioridad. Los autores de [7] proponen que sólo los nodos en la zona de encaminamiento difundan RREQ, con la intención de reducir la cantidad de los mismos. En [8] se presenta una idea similar, si bien en este caso los nodos se seleccionan en base a información de movilidad. La metodología descrita en este trabajo nunca se ha aplicado a AODV o DYMO. La funcionalidad de DYMOselfwd se presentó en [9] y en este trabajo se complementa con resultados de evaluación. Dichos resultados prueban las mejoras que se esperaban en la fase de diseño del algoritmo.

II. USO DE LA INFORMACIÓN DE POSICIÓN EN DYMO

El descubrimiento y la gestión de las rutas es una de las operaciones fundamentales del protocolo DYMO [1]. El descubrimiento de rutas se inicia cuando un nodo fuente (A)

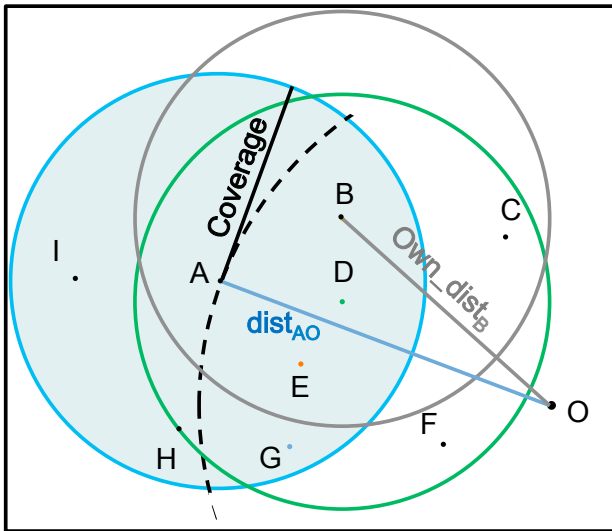


Fig. 1. Selección del mejor nodo que reenvía el RREQ enviado por A. El camino más corto (*own_dist*) lo tiene D.

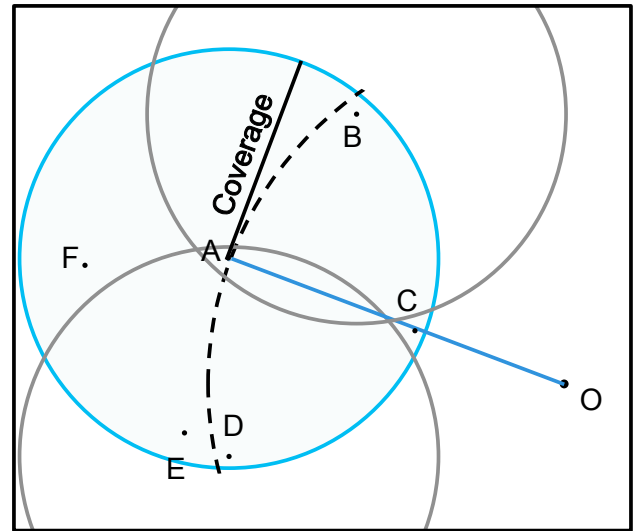


Fig. 2. Es posible que hasta tres nodos (B, C, D) puedan reenviar el RREQ a causa del problema del nodo oculto.

debe enviar datos a un nodo destino (*O*) y no está establecida una ruta entre ambos nodos. En tal caso el nodo *A* difunde un mensaje de control RREQ (Route Request). Cada nodo que recibe el RREQ se considera vecino de *A* y añade *A* en su tabla de encaminamiento con una distancia de 1 salto. Cada nodo vecino entonces comprueba si dispone de un camino hasta *O*. En caso afirmativo, se manda un RREP (Route Reply) a *A*. En caso contrario, el nodo en cuestión difunde un nuevo paquete RREQ, al que añade su propia dirección en el campo de información encargado de guardar el camino seguido por el paquete. De esta forma, el RREQ aumenta de tamaño en cada salto que se produce. Este procedimiento permite a los nodos intermedios conocer cuál debe ser el siguiente salto hacia los nodos que le preceden. Si uno de los nodos intermedios conoce el camino hacia *O*, este nodo envía un RREP a *A*, estableciendo así la ruta. De hecho, dicho mensaje establece el camino hacia *O* en todos los nodos entre él y el origen *A*.

A. Modificación del descubrimiento de ruta

Cuando *A* difunde el RREQ, DYMOselfwd incluye las posiciones de *A* y *O* en dicho paquete. Cada nodo intermedio sustituye la posición del nodo que ha enviado el RREQ por su propia posición. En el caso de que no haya RREP (el nodo intermedio no conoce la ruta al destino), se pretende que sólo uno de los nodos que ha recibido el RREQ realice la redifusión (es decir, enviar un nuevo RREQ con su propia posición). El procedimiento seguido para la selección del nodo encargado de realizar dicha redifusión queda detallado en la siguiente sección. Con esto se pretende una importante reducción de los mensajes RREQ necesarios para el establecimiento de la ruta.

Dada una determinada ruta, tanto DYMO como AODV requieren de la misma cantidad de RREQ para llevar a cabo el establecimiento del camino. Dicha cantidad es igual al número de nodos en la red dado que cada nodo que recibe el RREQ debe redifundirlo. Con DYMOselfwd dicha cantidad será proporcional al número de saltos entre fuente y destino. A pesar de que los paquetes RREQ son más pesados, por

contener éstos más información (las posiciones de dos nodos), se espera una importante reducción de la carga destinada al encaminamiento al reducir su número. Es posible que varios nodos redifundan el RREQ, en gran medida debido al problema del nodo oculto (véase la Fig. 2, donde en el caso de coberturas circulares el máximo es 3 nodos). Pese a todo, se espera que la reducción sea significativa y absorba el incremento en el número de RREQ que este escenario pueda suponer.

B. Selección del mejor nodo

El procedimiento seguido para seleccionar el mejor nodo en la ruta, es decir aquel nodo intermedio encargado de difundir en primer lugar el RREQ, bloqueando de esta forma las potenciales difusiones de otros nodos intermedios), sigue las ideas del BLR [4]. En BLR los paquetes se difunden hasta que llegan al destino sin necesidad de establecer ni mantener rutas. El mejor nodo se selecciona de una forma distribuida, aplicando diferentes retardos a los diferentes nodos candidatos. El BLR asigna el menor retardo al mejor nodo, entendiendo como mejor nodo aquel que dispone de la mejor posición dentro del área de reenvío. Dicho nodo es el encargado de redifundir el paquete. Esta idea se aplica de forma equivalente a DYMOselfwd. Cada nodo que recibe un RREQ relativo a una ruta desconocida (p. ej. *B* en la Fig. 1) sigue el siguiente procedimiento:

- Calcula la distancia ($dist_{AO}$) entre el nodo transmisor y el destino a partir de las posiciones incluidas en el RREQ.
- Calcula la distancia al destino desde su propia posición (Own_dist_B).
- Si la distancia desde el nodo intermedio al destino es más corta que la distancia entre origen y destino, el nodo se convierte en candidato a la redifusión. En caso contrario descarta el RREQ impidiendo la difusión en la dirección opuesta.

En este punto se utiliza un mecanismo de DFD (Dynamic Forward Delay) [4] para seleccionar el mejor nodo de una

manera distribuida. Dicho mecanismo se basa establecer de forma dinámica el retardo de redifusión del RREQ en cada nodo candidato, en función de las distancias previamente calculadas. Como consecuencia, algunos nodos candidatos no llegan a enviar un RREQ ya que durante el periodo de retardo, reciben otro RREQ y por tanto abortan el envío del suyo propio (interpretan que no son ellos los mejores candidatos). De esta forma, el mejor nodo será aquel que presente el menor retardo hacia el destino. La discusión sobre las posibles funciones DFD quedan fuera del alcance de este trabajo. Pese a lo indicado anteriormente, es posible que un nodo reciba varios RREQ, debido al problema del nodo oculto. En caso de que esto sucediera, el nodo tendrá en cuenta únicamente el primer RREQ, descartando los que se reciban con posterioridad a éste.

III. HERRAMIENTAS DE SIMULACIÓN Y ESCENARIOS

Para evaluar el funcionamiento del protocolo propuesto se ha utilizado el simulador Omnet++4.3 [10], junto al framework INET [11] para la construcción de los modelos de red. INET proporciona una completa implementación de varios protocolos de red y tecnologías inalámbricas especialmente útiles para este trabajo. De igual forma, permite el uso de múltiples modelos radio y de movilidad, algunos de ellos empleados en el presente estudio. El protocolo DYMOselfwd ha sido implementado a partir del módulo descrito en [12], tras evaluarse otros módulos que fueron descartados al no considerarse adecuados para este proyecto. Sobre el módulo finalmente seleccionado, se implementaron múltiples mejoras así como el código necesario para la implementación del propio DYMOselfwd.

Todas las simulaciones realizadas abarcan un periodo de 1.250 s. Durante este lapso de tiempo, el transmisor manda un paquete UDP cada 250 ms, resultando de esta forma en el envío de 5.000 paquetes en total. Cada experimento se lleva a cabo un total de 5 veces, empleándose el procedimiento de medias independientes para la provisión de resultados. En todos los casos los intervalos de confianza son mejores que el 95%. Se han propuesto dos escenarios con 9 y 13 nodos respectivamente. En cada uno de ellos se realizan pruebas con nodos estáticos (sin movilidad) y dinámicos (con movilidad). El patrón de movilidad empleado en los nodos se basa en el modelo Random Waypoint sin tiempos de pausa. El área considerada es de 150x150 m. La cobertura es de 71 m de manera que, incluso en el caso de 9 nodos, siempre hay al menos dos nodos al alcance en el escenario con nodos estáticos. La Tabla I proporciona un resumen de los principales parámetros empleados para la simulación, siguiendo las pautas presentes en otros trabajos similares [12].

La Fig. 3 muestra la distribución de los nodos en los escenarios estáticos y que sirve como condición inicial en los dinámicos. El escenario con 9 nodos (3x3) representa el caso de baja densidad y se considera una situación altamente restrictiva (poca conectividad). En este despliegue no hay más de dos nodos en la dirección del destino. Por su parte, el escenario con 13 nodos representa una densidad media. Ambos despliegues son denominados de baja densidad y densidad normal en el resto del trabajo. Siempre hay un nodo transmitiendo paquetes UDP al destino. La cobertura

TABLE I
CONFIGURACIÓN DE LOS PARÁMETROS

Velocidad	0 1 2 m/s
Protocolo	DYMO DYMOselfwd
Modelo de propagación	Two Ray Ground Model
Shadowing	0 dB
Tamaño del paquete	512 Bytes
Frecuencia portadora	2,4 GHz
Protocolo MAC	802.11g
Tasa de bit	54 Mbps

TABLE II
RESULTADOS CON NODOS ESTÁTICOS EN AMBOS ESCENARIOS.

	Baja densidad		Densidad normal	
	DYMO	DYMOselfwd	DYMO	DYMOselfwd
Troute [ms]	5,21	5,72	5,71	5,79
ColMac	15,8	20,2	30,2	0,0
MACPdu	20.253,0	20.262,4	20.351,6	20.176,0
REQtot	180,2	178,4	265,6	84,0

está representada por un círculo rojo y el mejor nodo para la retransmisión del RREQ se representa en gris. Las métricas obtenidas en las simulaciones y que son comparadas con el protocolo DYMO se explican a continuación:

- **Troute:** Tiempo necesario para establecer la ruta, es decir, hasta el instante en que el origen recibe el RREP.
- **ColMac:** Número de colisiones a nivel MAC.
- **MacPdu:** Número total de paquetes a nivel MAC enviados por todos los nodos.
- **REQtot:** Número total de peticiones RREQ enviadas por todos los nodos.

IV. RESULTADOS

A. Estático

La Tabla II muestra los resultados para este escenario, resultados que se pueden considerar representativos de lo que ocurre en una red con sensores estáticos. Para baja densidad DYMO funciona mejor que DYMOselfwd: las primeras tres métricas son mejores para DYMO mientras que la última es algo peor. Esto es debido a la simetría del despliegue. Dado que hay dos nodos exactamente a la misma distancia de la fuente (nodos 1 y 3 en la Fig. 3) ambos redifunden el RREQ al mismo tiempo sin escucharse entre sí con la consecuencia de que el número de RREQ es similar. Por eso, el número de nodos que envían el RREQ es igual en ambos protocolos y se obtienen resultados similares. Sin embargo, a causa de a la modificación introducida, el tiempo de establecimiento es superior debido al retardo introducido en los RREQ. Además, el retardo añadido aumenta el número de colisiones, que principalmente se dan en el nodo que está en el medio del escenario; en consecuencia, el número de MACPdu enviadas también aumenta.

Para confirmar esta hipótesis se han realizado pruebas alterando ligeramente la simetría mediante pequeños desplazamientos de los nodos. Los resultados obtenidos son muy similares tanto para DYMO como para DYMOselfwd: el

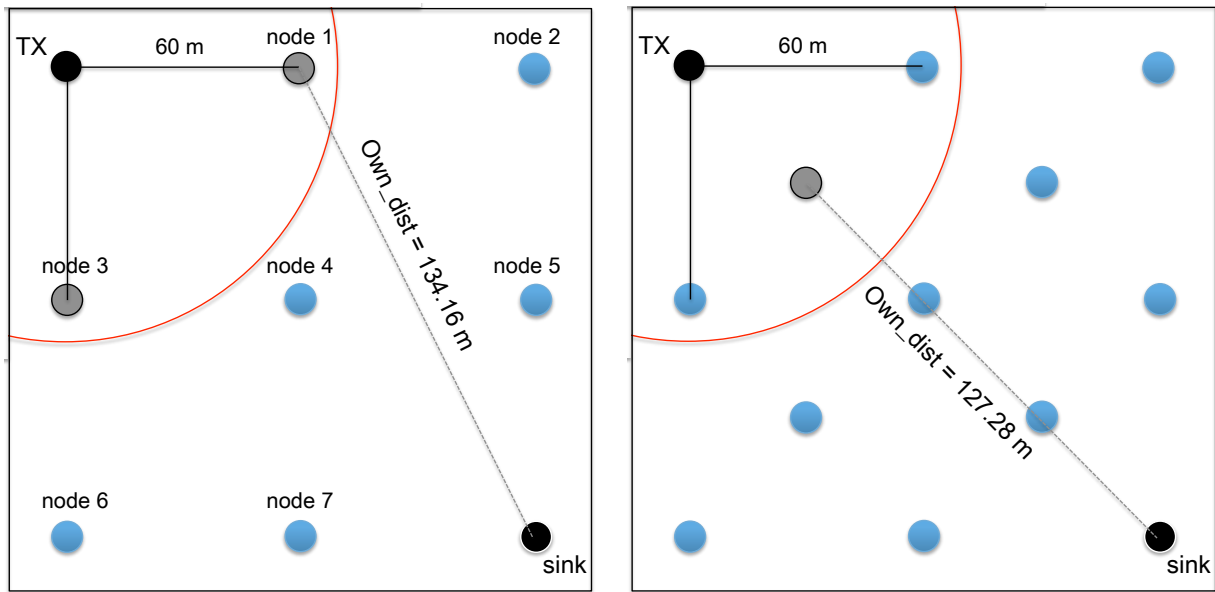


Fig. 3. Posición de los 9 nodos para el caso de baja densidad (a la izquierda) y de los 13 nodos para el caso de densidad normal (a la derecha). TX es el nodo que envía segmentos UDP al sink. Su área de cobertura está dibujada en rojo.

número de colisiones disminuye (ej. ColMax es 7 en ambos casos) y las prestaciones siguen siendo parecidas para ambos protocolos. Sin embargo, ahora el comportamiento de DYMOselfwd es ligeramente mejor que DYMO, de acuerdo a los valores reportados para Troute, MACPdu y REQtot, todos ellos inferiores en el caso del DYMOselfwd.

Para una densidad normal, las métricas arrojan mejores cifras en el caso del DYMOselfwd, con la salvedad del tiempo de establecimiento. Tanto las colisiones como el número de RREQ se reducen drásticamente mientras que el tiempo de establecimiento es prácticamente el mismo.

Las siguientes secciones muestran los resultados en diferentes situaciones en cuanto a movilidad. Para facilitar la comparación, los resultados se presentan en la Fig. 4. En la Fig. 4(a) se muestran los resultados para Troute; en la Fig. 4(b), 4(c) y 4(d) los de ColMAC, MACPdu y REQtot, respectivamente. Los valores en el caso estático se han representado con línea negra, los de velocidad media con línea discontinua azul y finalmente los de velocidad alta con una línea verde de rayas y puntos. Los resultados de DYMO emplean un círculo como marca mientras que los de DYMOselfwd realizan lo propio con un asterisco.

B. Velocidad media

Este escenario representa lo que ocurre en una MANET con velocidad de 1 m/s (ej. peatones). Las conclusiones son similares a las ya presentadas en el caso estático: DYMOselfwd mejora solamente a partir de un cierto grado de conectividad. En este caso las consecuencias de la baja conectividad de los nodos y simetría del escenario (9 nodos) son peores que para el caso estático en lo referente al tiempo de establecimiento. Ello se debe a que en el caso estático el nodo fuente siempre disponía de dos nodos a su alcance, mientras que ahora, dependiendo de la movilidad de los nodos, el nodo fuente puede tener uno o ningún nodo a su alcance. Con baja o muy baja conectividad una estrategia de difusión completa es

sin duda más rápida pero también implica mayor sobrecarga: el número de RREQ y de colisiones es siempre inferior con DYMOselfwd.

C. Velocidad alta

En este caso, todos los resultados son peores que para los casos estático y de movilidad media. Es bien conocido el coste de la movilidad en redes inalámbricas, en las que habitualmente, una mayor velocidad de desplazamiento en los nodos se traduce en una mayor carga de gestión en la red.

Por lo demás y aparte del retardo ligeramente superior (esperado debido a la introducción de retardo en los envíos de RREQ), el funcionamiento de DYMOselfwd es mejor en las áreas previstas, presentando una reducción del número de colisiones y de los RREQ totales enviados en la red.

V. CONCLUSIONES

La investigación sobre algoritmos de encaminamiento en redes MANET es un tema todavía abierto. Los mecanismos más convencionales presentan problemas de escalabilidad debido a la elevada difusión que realizan de los paquetes de control. El uso del conocimiento de la posición de los nodos en redes cuyos dispositivos integran capacidad de localización resulta clave para reducir la necesidad de difusión de paquetes de control, lo cual resulta esencial de cara a mantener la escalabilidad de la red.

El algoritmo propuesto en este trabajo y del que se presenta una primera evaluación se basa en los principios mencionados. Toma como base el protocolo DYMO, al que se le aplican procedimientos inspirados en el BLR, todos ellos basados en el uso de la posición. De esta forma, al encaminamiento DYMO se introducen retardos en función de las distancias al origen, tal y como se describe en el BLR. La consecuencia inmediata es una drástica reducción de los paquetes RREQ necesarios, debido a la inhibición de los nodos más lejanos en la ruta hacia el destino, en favor de aquellos más próximos. Todo

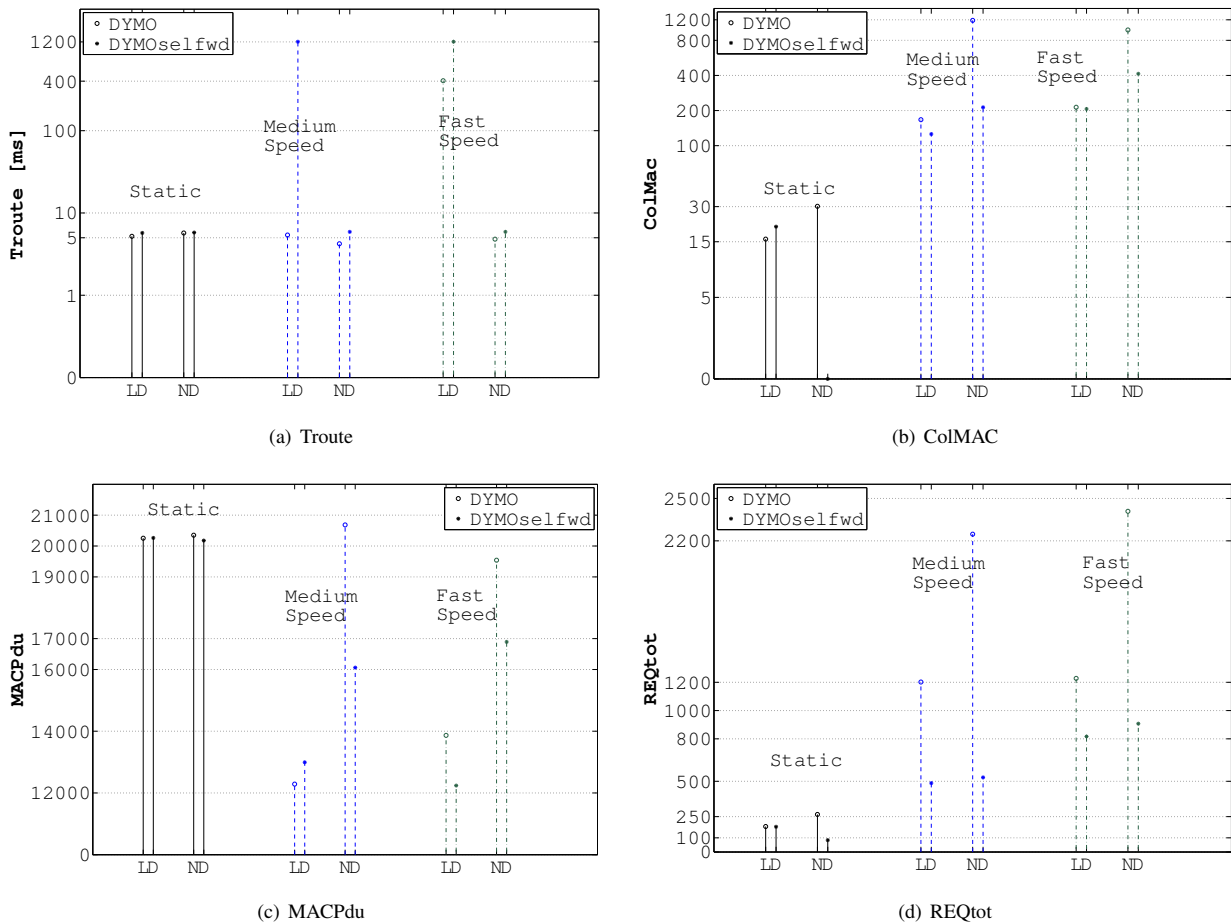


Fig. 4. Resultados para DYMO (o) y DYMOselfwd (*) para los casos estático (línea negra), velocidad media (línea discontinua azul) y velocidad alta (línea verde de rayas y puntos). Por cada uno, se muestra los casos de baja densidad (LD) y densidad normal (ND).

ello redundando en una mejora de la capacidad de la red para transmitir información, mejorando de manera significativa la escalabilidad de la red. Como contrapartida, el DYMOselfwd introduce un ligero aumento del retardo.

Como futuras líneas de investigación de este trabajo se considera valorar el impacto que puedan tener las imprecisiones de la posición sobre el funcionamiento del algoritmo. De igual forma, se pretende evaluar el impacto del shadowing y de modelos de movilidad diferentes al Random Waypoint.

AGRADECIMIENTOS

Jordi Ruiz Paños y Raúl Ruiz Díaz colaboraron en la implementación y procesamiento de los datos de las simulaciones presentadas en este trabajo. Los autores quieren agradecer su valiosa ayuda.

La investigación ha sido financiada a través de CICYT y FEDER en el marco del proyecto TEC2013-48099-C2-1-P.

REFERENCES

[1] C. Perkins, I. Chakeres, Dynamic MANET On-demand (AODVv2) Routing, IETF Internet Draft (Standards Tracks, work in progress), 2012. <<http://tools.ietf.org/html/draft-ietf-manet-dymo-22#page-33>>.
 [2] Y. Ko, N.H. Vaidya, Location-Aided Routing (LAR) in mobile ad hoc networks, in: Proc. ACM MobiCom, 1998, pp. 66-75.

[3] V. Giruka, M. Singhal, A self-healing on-demand geographic path routing protocol for mobile ad-hoc networks, Elsevier Ad Hoc Networks, 5 (7), (2007) 1113-1128.
 [4] M. Heissenbüttel, T. Braun, T. Bernoulli, M. Wälchli, BLR: Beacon-Less Routing Algorithm for Mobile Ad-Hoc Networks, Elsevier Computer Communications Journal (Special Issue), 27, (2003) 1076-1086.
 [5] C. H. Chou, K.F. Ssu, H.C. Jiau, Dynamic route maintenance for geographic forwarding in mobile ad hoc networks, Elsevier Computer Networks, 52(2), (2008) 418-431.
 [6] K. Endo, Y. Inoue, Y. Takahashi, Performance modeling of beaconless forwarding strategies in multi-hop wireless networks, Elsevier Computer Communications, 35 (1), (2012) 120-128.
 [7] R. Reno, Enhanced AODV for directional flooding using coordinate system, in: Proc. International Conference on Networking and Information Technology (ICNIT), 2010, pp. 329-332.
 [8] Y. Khamayseh, O.M. Darwish, S.A. Wedian, MA-AODV: Mobility Aware Routing Protocols for Mobile Ad Hoc Networks, in: Proc. Fourth International Conference on Systems and Networks Communications, 2009, pp. 25-29.
 [9] E. Zola, F. Barcelo-Arroyo, DYMO Routing Protocol with Knowledge of Nodes' Position, in: Proc. Joint ERCIM eMobility and MobiSense Workshop, 2012, pp. 9-14.
 [10] Omnet++ release 4.3. <<http://www.omnetpp.org/>>.
 [11] The INET framework. <<http://inet.omnetpp.org/>>.
 [12] C. Sommer, I. Dietrich, F. Dressler, Simulation of Ad Hoc Routing Protocols using OMNeT++, Springer Mob. Netw. Appl, 15 (6), (2010) 786-801.

Uso de OpenFlow para la gestión de caminos redundantes en redes inalámbricas

Elías Molina, Armando Astarloa, Eduardo Jacob.
Departamento de Ingeniería de Comunicaciones,
Universidad del País Vasco (UPV/EHU)
Alameda Urquijo s/n. 48013, Bilbao - Bizkaia
{elias.molina, armando.astarloa, eduardo.jacob} ehu.es

Resumen—El despliegue de redes inalámbricas en entornos industriales críticos debe garantizar una alta disponibilidad de los servicios, para lo cual se hace indispensable el uso de técnicas de redundancia. Con el objetivo de minimizar a cero el tiempo de recuperación ante errores, en este artículo se propone el uso de la tecnología Parallel Redundancy Protocol (PRP), el cual emplea el concepto de redundancia activa, en infraestructuras Wireless Local Area Networks (WLAN) que estén implementadas bajo el paradigma Software-Defined Networking (SDN). Se analizan las ventajas obtenidas al controlar los recursos redundantes a través del protocolo OpenFlow, con el que se establecen los caminos redundantes para aquellos servicios que así lo requieran. Lo que supone una adecuada gestión de la redundancia, esencial para maximizar el grado de disponibilidad y resiliencia de una red. El uso de herramientas de emulación ha permitido evaluar las mejoras obtenidas en cuanto a disponibilidad de las comunicaciones en situaciones de pérdidas y movilidad.

Palabras Clave—IEC 62439, OpenFlow, Parallel Redundancy Protocol, Redundancia, Software-Defined Networking

I. INTRODUCCIÓN

Tradicionalmente, los escenarios industriales han empleado comunicaciones cableadas para la transmisión de información crítica. Sin embargo, tal y como se identifica en [1], surge la necesidad de desplegar sistemas de monitorización y control inalámbricos que reduzcan el coste con respecto a la instalación y mantenimiento de comunicaciones alámbricas. Además, tales escenarios demandan con más frecuencia infraestructuras de redes de datos que permitan la interconexión de nodos móviles, así como un rápido despliegue. Para lo cual, las redes inalámbricas resultan una solución adecuada.

A pesar de que las redes inalámbricas adolecen de un medio de transmisión poco fiable debido a su comportamiento no determinista, tales infraestructuras han de asegurar la continuidad del servicio ofrecido, de forma que no se comprometa la disponibilidad ni el rendimiento. Para alcanzar dicho objetivo, las redes deben disponer de elementos redundantes que eviten puntos únicos de fallo, siendo capaces de restaurar las comunicaciones ante fallos e interferencias en la red. Concretamente, en este artículo se analizan las redes basadas en el estándar IEEE 802.11, como caso destacado en el ámbito de las redes locales.

Son numerosos los entornos de comunicaciones críticas que requieren latencias y pérdidas de información mínimas. Algunos de los cuales se encuentran en:

- *Transporte*: soluciones altamente fiables deben ser implementadas en diferentes sectores, tales como sistemas de control de tráfico, redes vehiculares o aviónica.

- *Sistemas de automatización*: por ejemplo, las *smart grids* tienen que garantizar el suministro de energía en las subestaciones, y líneas de transmisión y distribución eléctricas. De acuerdo a las directrices para la ciberseguridad en las *smart grids* dadas por el *National Institute of Standards and Technology* (NIST), la disponibilidad de las infraestructuras es considerado el requisito de seguridad más crítico [2]. Un enfoque similar se encuentra en las redes de control industrial.

Del último caso de uso crítico, sirva como ejemplo el estándar IEC 61850, orientado a las comunicaciones de datos en las instalaciones de suministro eléctrico. Dicho estándar detalla los requerimientos de las redes locales para cada tipo de servicio que tiene lugar en las subestaciones eléctricas. En particular, la especificación IEC 61850-5 [3] detalla los tiempos de recuperación permitidos ante errores, exigiendo una total continuidad para los servicios más críticos. Por ejemplo, los mensajes de datos asociados a la protección de líneas y transformadores eléctricos, lo cual es desempeñado por los diferentes relés de medición y protección, requieren un tiempo de recuperación nulo, de forma que se asegure que no se produzca pérdida de información alguna.

Para la consecución de tales requisitos, el IEC 61850 prescribe la utilización de los mecanismos de redundancia definidos en el estándar IEC 62439-3 [4], entre los que se encuentra el Parallel Redundancy Protocol (PRP). Al contrario de lo que ocurre con sistemas de redundancia pasiva, en los que las conexiones de red conmutan hacia recursos de reserva cuando sea necesario, PRP es un protocolo de redundancia activa, en el que los nodos finales transmiten múltiples copias de datos a lo largo de dos rutas diferentes de forma simultánea. Este enfoque evita tiempos de interrupción, asegurando que no se pierden mensajes en caso de fallo en una de las rutas.

Teniendo en cuenta que PRP no especifica la forma de establecer las rutas en la red de datos, en este artículo se propone el uso de OpenFlow para la configuración de flujos redundantes en entornos inalámbricos. Con OpenFlow los planos de datos y control están desacoplados, acorde al paradigma SDN [5]. Así, la solución propuesta se beneficia de utilizar un agente externo que posee una visión global del estado de la red, permitiendo un control de los recursos de red flexible para diferentes servicios.

El resto del artículo se estructura como sigue: en la Sección II se destacan las características del PRP; la Sección III analiza las oportunidades de las redes inalámbricas en entornos críticos. La especificación OpenFlow es descrita en la Sección IV, mientras que el diseño propuesto se presenta

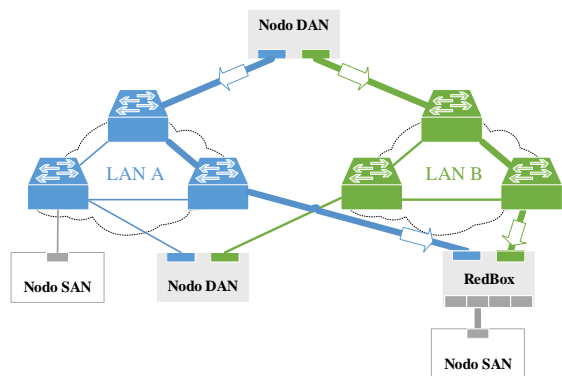


Fig. 1: Esquema de red con nodos PRP.

en la Sección V. La implementación técnica y resultados son presentados en la Sección VI. Finalmente, la Sección VII contiene las conclusiones y líneas futuras del trabajo.

II. PARALLEL REDUNDANCY PROTOCOL

La tecnología PRP es implementada en los nodos finales, denominados Double Attached Nodes (DANs, ver Figura 1), los cuales se conectan a dos redes LAN independientes a través de sendas interfaces con la misma dirección MAC. Dichas redes pueden tener diferentes topologías y prestaciones, siendo en cualquier caso agnósticas al protocolo PRP y pudiendo acoger a nodos no compatibles con PRP (Single Attached Nodes, SANs).

Para implementar redundancia activa, un DAN incluye un elemento denominado Link Redundancy Entity (LRE), encargado de gestionar la duplicación de tramas de nivel 2 de forma transparente para las capas superiores. El proceso de funcionamiento es el siguiente:

- 1) En el emisor, cuando el LRE recibe un mensaje de capas superiores, este crea dos tramas de nivel 2 a las que les añade información en forma de tráiler, denominado Redundancy Control Trailer (RCT), y calcula un nuevo *checksum*. El RTC permitirá en recepción detectar la información duplicada.
- 2) El LRE emite las tramas a través de los dos puertos del nodo, las cuales atravesarán dos redes independientes.
- 3) El DAN receptor espera recibir tráfico en ambos puertos, siempre y cuando las dos redes no fallen simultáneamente. Así, con el objetivo de transmitir a las capas superiores únicamente la primera trama recibida, el LRE destinatario implementa un registro que permite reconocer y descartar tramas duplicadas.

Además, los nodos PRP generan tráfico *multicast* de supervisión específico (Ethertype 0x88FB) que permite monitorizar el estado de las redes y la existencia de otros dispositivos PRP. Adicionalmente, el estándar define el uso de *proxies* PRP, denominados Redundancy Boxes (RedBoxes), a los cuales pueden conectarse nodos que no sean compatibles con PRP con el fin de proteger sus comunicaciones, como si de nodos PRP se tratase.

A pesar de que PRP estaba orientado inicialmente para ser utilizado en redes Ethernet cableadas, en este artículo se estudia la combinación de PRP en redes IEEE 802.11 con vistas a utilizar dispositivos de monitorización y control con interfaces inalámbricas redundantes.

III. TECNOLOGÍA INALÁMBRICA EN REDES INDUSTRIALES

En [6] se analizaron cómo soluciones inalámbricas maduras, como las definidas en los estándares IEEE 802.11, 802.15.1 o 802.15.4, pueden ser empleadas en aplicaciones industriales, a través de diseños de red robustos frente a canales inalámbricos propensos a errores.

En el caso de las mencionadas *smart grids*, diferentes artículos desglosan las oportunidades y retos que suponen las redes de sensores inalámbricas. Concretamente, en [1] los retos técnicos descritos, tales como asegurar una determinada latencia o calidad de servicio, están asociados a los cambios en la topología y conectividad debidos a las condiciones del medio físico, cuyas pérdidas de propagación son obtenidas por los autores en subestaciones eléctricas. De igual modo, los autores de [7] analizan las latencias obtenidas para mensajes definidos por el IEC 61850 en redes IEEE 802.11g, bajo niveles de interferencias previamente caracterizados en diferentes subestaciones y escenarios.

A. PRP-WLAN

Con la proliferación del uso de redes inalámbricas para la transmisión de datos críticos, se han elaborado recientes investigaciones que proponen la utilización de PRP en estos entornos, con el objetivo de aumentar la resiliencia de la red.

Para mejorar la disponibilidad de los servicios de medición en subestaciones basadas en el estándar IEC 61850, los autores de [8] estudian el uso de dispositivos con interfaces cableada e inalámbrica, operando de forma similar a PRP; es decir, recibiendo datos duplicados a través de ambas conexiones. Es en [9] donde se propone la conexión de dos nodos PRP mediante dos canales 802.11 independientes que operan como enlaces punto a punto. Como resultado, se obtiene una comunicación entre emisor y receptor con una menor tasa de errores, latencia y *jitter*. Con el mismo enfoque, en [10] se optimiza el envío de tramas redundantes con la incorporación de mensajes de control, los cuales evitan la transmisión sobre canales inalámbricos que están sufriendo interferencias y pérdidas.

Estas propuestas demuestran las ventajas que supone combinar PRP y WLANs en entornos con exigentes requisitos de disponibilidad. Sin embargo, no se han analizado conexiones punto a multipunto ni escenarios de movilidad. Además, ninguna de las propuestas anteriores está enfocada en la gestión de la red; mientras que, en este estudio se propone OpenFlow como mecanismo de control.

IV. OPENFLOW

En esta sección se presenta una descripción de OpenFlow, estandarizado y promovido por la *Open Networking Foundation*. Mediante el protocolo OpenFlow, un controlador puede decidir el encaminamiento de los diferentes flujos que llegan a un *switch*. Para lo cual, añade entradas en las tablas de encaminamiento del *switch*, definidas a través de diferentes campos (puerto de entrada y cabeceras de paquetes), prioridades e instrucciones asociadas a cada flujo de entrada [11].

El establecimiento de las reglas de encaminamiento puede realizarse de manera reactiva, en respuesta a consultas realizadas por los *switches* ante la recepción de nuevos

flujos, o proactiva, a través de la preinstalación de reglas. El comportamiento proactivo es requerido para aquellas comunicaciones *time-sensitive*.

En cuanto a la capacidad de recuperación ante errores de una red OpenFlow, en [12] se recogen diferentes estudios que analizan cómo la incorporación de múltiples controladores, reglas de encaminamiento de respaldo, así como la detección de fallos delegada en los propios dispositivos de red, permiten implementar redes altamente robustas.

La utilización de tecnologías SDN para la gestión de infraestructuras críticas permite beneficiarse del concepto de programabilidad de la red. En el trabajo [13], el paradigma SDN ha sido propuesto para implementar prestaciones tales como filtrado de tráfico, balanceo de carga o calidad de servicio en sistemas basados en el estándar IEC 61850.

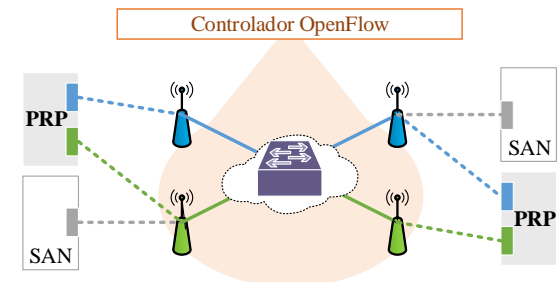
En relación con el objeto de estudio, en el proyecto OpenRoads [14], OpenFlow es usado para establecer activamente enlaces inalámbricos redundantes, ya sean bajo tecnología WiFi como WiMAX. Así, los autores consiguen una comunicación sin interrupciones en situaciones de itinerancia. Sin embargo, la detección y supresión de tráfico duplicado por parte del nodo receptor es realizada con un módulo del kernel de Linux (*nf_mobility*). Dicho módulo establece un número de secuencia por conexión (protocolo, direcciones IP origen y destino y puertos de la capa de transporte). Este método, al contrario que PRP, no es compatible con aquellos protocolos que envíen los datos directamente en capas inferiores; tal y como ocurre, por ejemplo, en los servicios definidos por el IEC 61850: GOOSE y SV (ambos transportan información crítica y son mapeados directamente sobre tramas de nivel 2).

La combinación de los protocolos OpenFlow y PRP ha sido propuesta en un trabajo previo [15]. Dicho artículo tiene como objetivo mejorar las prestaciones que ofrecen las redes tradicionales a las que se conectan los nodos PRP, las cuales están generalmente basadas en protocolos de *spanning tree*, como por ejemplo el Rapid Spanning Tree Protocol (RSTP, IEEE 802.1D). Por consiguiente, se consiguen ventajas tales como soportar múltiples fallos en la red mediante el establecimiento de más de dos caminos en una misma red, lo cual no es factible bajo RSTP.

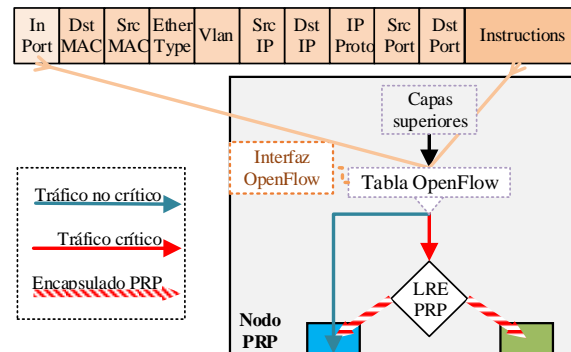
Sin embargo, aún no han sido estudiados los beneficios que esta combinación puede ofrecer en entornos inalámbricos, en los que las conexiones son más susceptibles a pérdidas e interferencias.

V. PROPUESTA PARA LA GESTIÓN DE CAMINOS INALÁMBRICOS REDUNDANTES

En la solución propuesta, los dispositivos finales se conectan a uno o varios puntos de acceso de forma redundante, siendo un controlador OpenFlow el que establezca las reglas de encaminamiento de tráfico en los dispositivos de distribución (Figura 2a). Además, los nodos finales incorporan también un procesamiento de tráfico basado en OpenFlow, pudiendo incorporar reglas que les permita distinguir diferentes flujos. De esta forma, la provisión de rutas será establecida extremo a extremo en función de la criticidad del tráfico. Es decir, entre nodos emisor y receptor se establecerán rutas paralelas para aquellas tramas que requieran una continuidad total del servicio, pero no para el



(a) Diagrama de red inalámbrica con nodos PRP bajo control OpenFlow.



(b) Esquema de un nodo PRP con filtrado de tráfico en base a OpenFlow.

Fig. 2: Combinación de las tecnologías PRP y OpenFlow en entornos inalámbricos.

tráfico que no sea considerado crítico. Por consiguiente, el plano de control de la red es consciente de las necesidades de los nodos PRP, al contrario de lo que sucede en las redes actuales.

El controlador OpenFlow establece las reglas que permiten que dos nodos PRP se comuniquen entre sí. Si se desea una mayor robustez, las rutas establecidas deberán ser disjuntas, tal y como se describe en el estándar IEC 62439-3. Aunque, como se estudia en [15], un control basado en OpenFlow permite configuraciones flexibles en las que las redes a las que se conectan los nodos PRP puedan compartir recursos.

A. Políticas de filtrado de tráfico en nodos PRP

Esta propuesta habilita la posibilidad de utilizar el protocolo PRP según sea necesario, lo cual supone un mejor aprovechamiento de los recursos disponibles. El control del tráfico en los nodos finales es realizado en base a políticas de prioridad, es decir, los servicios críticos serán protegidos con el formato de trama propio de PRP; mientras que, los servicios que toleren cierta interrupción en la comunicación serán transmitidos a través de una de las interfaces disponibles, sin encapsulamiento PRP. Por ejemplo, algunas de las entradas configuradas en un nodo PRP podrían ser las siguientes:

```

{"Tabla de flujos": {
  "id": "prp1",
  "entradas": [
    {"dl_type": "0x88b8", "actions": "output:lre_prp"},
    {"in_port": "prp1-eth0", "dl_type": "0x88b8",
     "actions": "output:lre_prp"},
    {"in_port": "prp1-eth1", "dl_type": "0x88b8",
     "actions": "output:lre_prp"},
    {"dl_type": "0x0800", "nw_dst": "192.168.60.2",
     "nw_proto": "17", "tp_dst": "161",
     "actions": "output:prp1-eth0"}
  ]
}
    
```

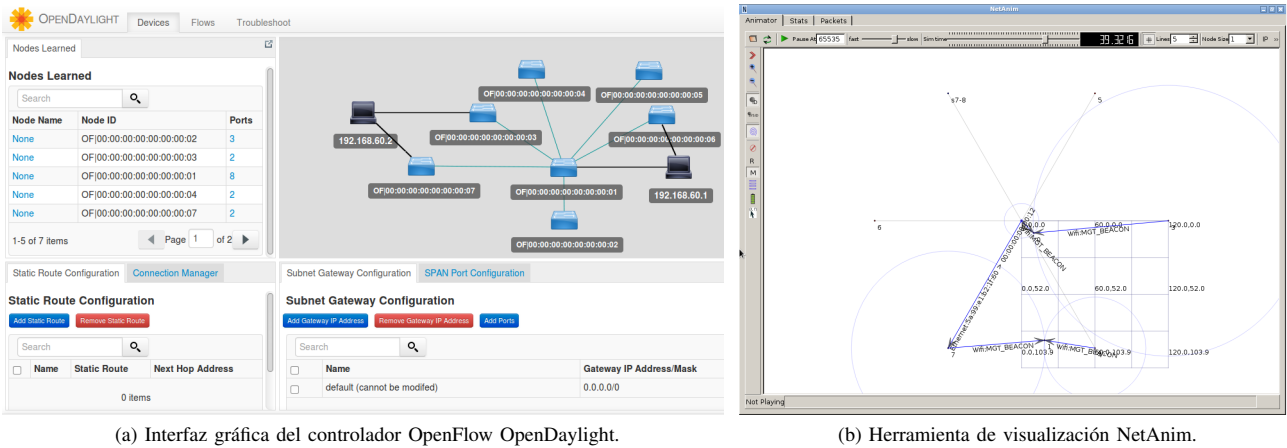


Fig. 3: Herramientas utilizadas en el proceso de emulación.

En concreto, la primera regla establece que todo el tráfico con el Ethertype *0x88b8*, correspondiente a tráfico GOOSE, salga por el puerto interno *lre_prp*, el cual añadirá el tráiler RCT a las tramas y las sacará por las interfaces físicas *prp1-eth0* y *prp1-eth1* (las dos reglas siguientes permiten seguir el camino inverso). En cambio, la última regla establece que el tráfico IP, con dirección destino *192.168.60.2*, UDP y puerto destino *161* salga directamente por el puerto *prp1-eth0*.

El flujo de datos seguido en los nodos descritos es representado en la Figura 2b, junto con los campos definidos en OpenFlow [11] que pueden servir para el filtrado. Estas políticas son instaladas con mensajes OpenFlow. Sin embargo, los nodos finales quedan fuera de la visión del controlador, destinado al control de los nodos de distribución.

VI. SIMULACIÓN

A. Implementación técnica

En cuanto al desarrollo técnico, la solución utiliza diferentes proyectos software de código abierto.

- Como controlador OpenFlow se ha utilizado OpenDaylight [16]. Este provee un plano de control completo, responsable de tareas como el descubrimiento de dispositivos y topología de red, así como el cálculo de rutas y la prevención de bucles.
- Como elemento de distribución se utiliza el *switch* virtual Open vSwitch (OVS) [17], el cual soporta el protocolo OpenFlow. Además, OVS es incorporado también en los nodos PRP para realizar el filtrado de flujos.
- Los nodos finales utilizan una implementación software del procesamiento PRP, que se ejecuta en espacio de usuario bajo el sistema operativo Linux.
- Para conformar la emulación de la red, todos los elementos son conectados mediante OpenNet [18], una herramienta de simulación reciente basada en Mininet [19]. OpenNet permite comunicar interfaces creadas en Mininet con enlaces modelados con el simulador de red ns-3. Como consecuencia, pueden evaluarse conexiones de red inalámbricas y con nodos móviles, lo cual no es posible con Mininet.

En la Figura 3 se representan dos interfaces gráficas utilizadas en el proceso de emulación: *a*) la interfaz de gestión de OpenDaylight, en la que en tiempo real puede

obtenerse la topología de una red, así como información sobre los dispositivos de red y reglas de encaminamiento instaladas; *b*) la herramienta NetAnim de ns-3, que permite un análisis posterior en el que visualizar las tramas transmitidas, así como el movimiento de los nodos.

En cuanto a la configuración del escenario de pruebas, la topología y condiciones de red utilizadas en los resultados aquí presentados son:

- Uso del estándar de transmisión 802.11g, el cual opera en la banda de 2.4 GHz y alcanza una velocidad promedio real de transferencia de 22,0 Mbit/s.
- La topología de red opera en modo infraestructura, en la que los clientes se conectan a un punto de acceso en una topología en estrella, y los puntos de acceso están cableados a la red de distribución mediante interfaces 100 Mbit/s *full-duplex*.
- El tráfico de control OpenFlow es transmitido a través de recursos independientes del plano de datos (configuración *fuera de banda*), siendo también posible una configuración *en banda*.

A continuación, se describen dos casos de estudio que ponen de relieve el aumento en la disponibilidad de las comunicaciones al establecer conexiones paralelas. En concreto, se analiza una situación en la que los enlaces inalámbricos no ofrecen total fiabilidad, así como una situación de movilidad entre diferentes puntos de acceso.

B. Mayor robustez frente a pérdidas

Como se ha descrito anteriormente, PRP evita las interrupciones en caso de que una de las dos rutas establecidas falle. Esto se corresponde con la disponibilidad global de sistemas paralelos frente a sistemas conectados en serie:

- La disponibilidad ($A \in [0, 1]$) de un sistema en serie es igual al producto de la de sus subsistemas (N),

$$A_s = \prod_{i=1}^N A_i$$

- Mientras que, un sistema redundante en paralelo ve incrementada su disponibilidad con cada uno de sus sistemas,

$$A_p = 1 - \prod_{i=1}^N (1 - A_i)$$

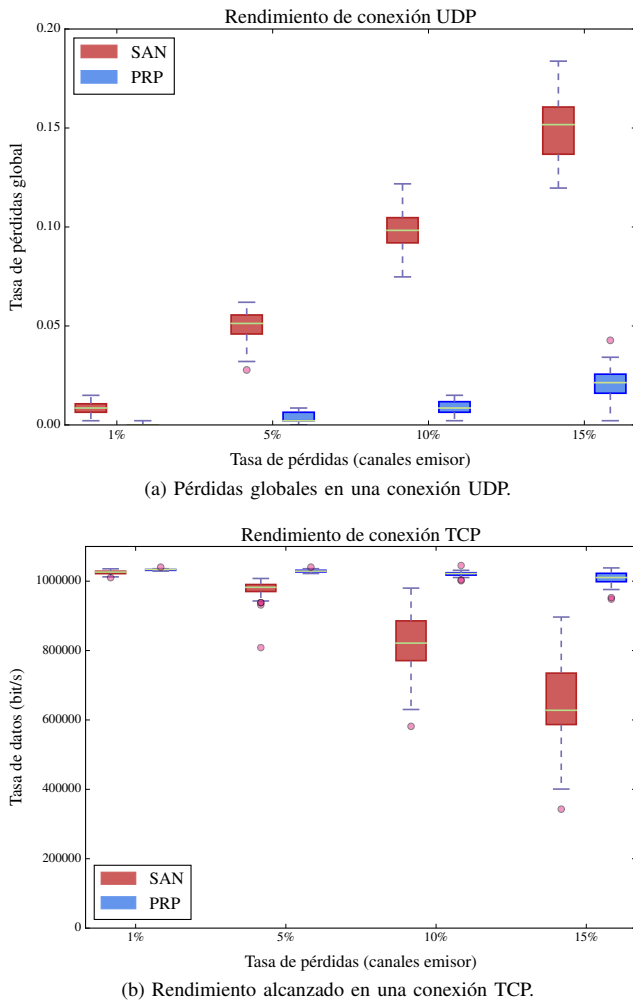


Fig. 4: Efecto de disponer enlaces inalámbricos con pérdidas en conexiones fiables o no fiables.

- La indisponibilidad ($U = 1 - A$, $U \in [0, 1]$) de los subsistemas de red puede asociarse directamente con la tasa de pérdida de paquetes global ocasionada. Por tanto, un sistema como PRP, que duplica la cadena de subsistemas, reduce dicha tasa con respecto a un sistema no redundante. Para una cadena con indisponibilidad q_1 , siendo q_2 la de la cadena redundante,

$$U_{no\ redundante} = q_1 \geq U_{paralelo} = q_1 * q_2$$

Para mostrar la robustez frente a pérdidas en ambas rutas, se ha estudiado la disponibilidad de un servicio cuando los dos enlaces inalámbricos a los que está conectado un nodo PRP tienen la misma probabilidad de pérdida de paquetes, siendo estadísticamente independientes entre sí. En la Figura 4a se representa la disponibilidad frente a la tasa de pérdidas en los casos en los que se dispone de un único enlace y en un funcionamiento basado en PRP. Las pruebas han sido realizadas con tráfico UDP (transporte no fiable), cuando en los canales a los que se conecta el emisor se producen pérdidas con diferente probabilidad.

Para evidenciar lo que supone este incremento en la disponibilidad sobre servicios que funcionan sobre TCP (transporte fiable), en la Figura 4b se muestra la tasa de

datos alcanzada en una conexión TCP limitada en origen a 1.05 Mbit/s. Los gráficos de caja (*boxplots*, que incluyen valor mínimo, primer cuartil, mediana, tercer cuartil y valor máximo) representados corresponden a 30 muestras de las que se recoge información durante 10 segundos.

C. Continuidad del servicio en itinerancia

En este caso de uso se estudia cómo, al establecer conexiones simultáneas con diferentes puntos de acceso, la comunicación permanece inalterada cuando un nodo pierde la cobertura con uno de los puntos de acceso. Teniendo en cuenta que las tecnologías SDN/OpenFlow permiten optimizar la provisión de recursos ante la predicción de cambios de puntos de acceso por parte de los nodos finales (*handover*), tal y como se analiza en [20], se han comparado también las pérdidas producidas en un proceso de *handover* cuando los puntos de acceso están transmitiendo datos de forma preventiva antes del cambio en la conexión AP-nodo final. Sin embargo, el proceso de escaneo y conexión supone en cualquier caso una interrupción en la comunicación que no ocurre al utilizar PRP, como se muestra en la Figura 5b. En particular, en esta figura se representan tres casos (Figura 5a) en los que un nodo que recibe tráfico UDP se mueve entre dos puntos de acceso que operan en diferentes canales 802.11g:

- 1) Nodo SAN con única interfaz:
 - a) El controlador OpenFlow instala reglas en los elementos de red propias de un funcionamiento de aprendizaje tradicional (*learning switch*). Esto implica que se incurra en un mayor tiempo de recuperación, pues el controlador debe reaccionar ante la transmisión de datos en un enlace diferente con la instalación de nuevas reglas y modificación o borrado de las anteriores.
 - b) El controlador instala reglas en los elementos de red de forma tal que el tráfico es duplicado en los puntos de acceso con vistas a reducir el periodo de interrupción.
- 2) Nodo con dos interfaces: se conecta a los dos puntos de acceso entre los que se mueve, por lo que permanece conectado en todo momento.

Sirvan estos tres ejemplos también para representar como un mismo nodo puede al mismo tiempo generar tráfico encapsulado o no sobre PRP. Esto convendrá según la criticidad del tráfico y el estado de la red.

VII. CONCLUSIONES Y LÍNEAS FUTURAS

Son numerosas las aplicaciones que requieren topologías redundantes que ofrezcan alta disponibilidad, al mismo tiempo que encuentran en las redes inalámbricas numerosas ventajas, tales como el rápido despliegue o la habilitación de nodos móviles. Con el propósito de incrementar la disponibilidad, este artículo estudia las prestaciones obtenidas en redes 802.11, en las que los nodos están conectados redundantemente y cuyo tráfico está controlado a través de OpenFlow. Además, los nodos PRP permiten incorporar reglas de encaminamiento OpenFlow, pudiendo filtrar aquellos flujos que requieran protección activa de los que no. En particular, mediante emulación se analiza la continuidad de los servicios críticos ante posibles fallos en la red, así como en situaciones de itinerancia de nodos finales entre puntos de acceso.

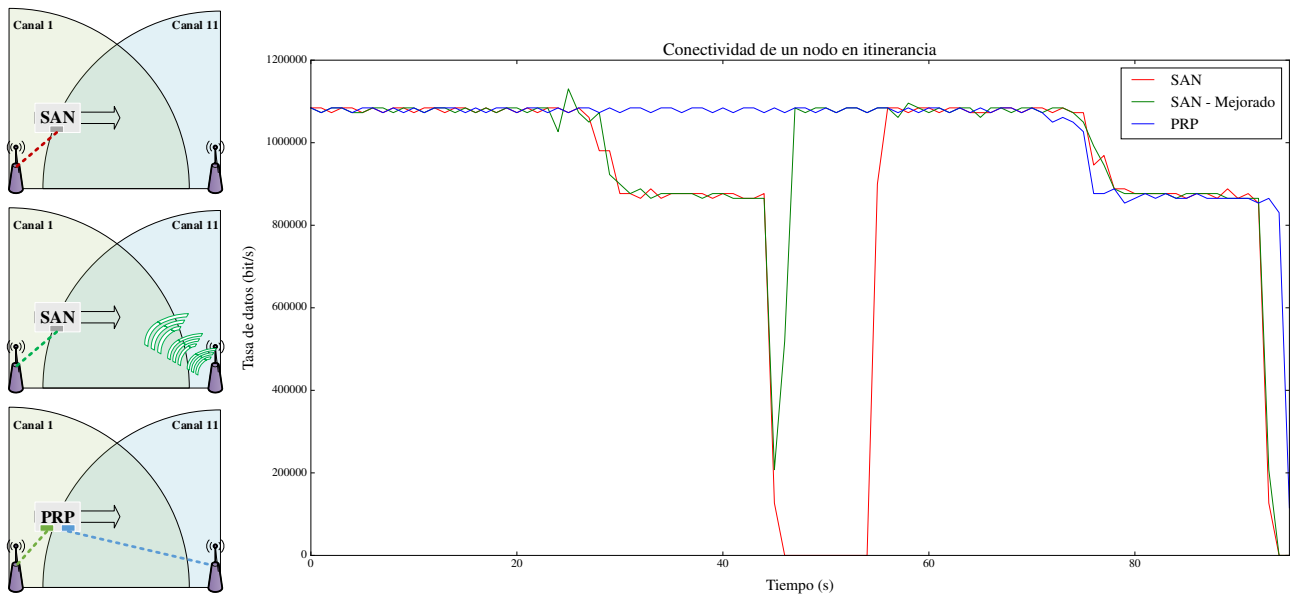


Fig. 5: Continuidad en la conectividad en un nodo que se mueve entre puntos de acceso.

Con respecto a las líneas futuras, consideramos interesante analizar el rendimiento obtenido en redes mallas IEEE 802.11s, así como bajo el estándar 802.11n.

AGRADECIMIENTOS

El trabajo descrito ha sido posible gracias al programa ZABALDUZ, habiéndose producido en la Unidad de Formación e Investigación UFI11/16 financiada por la Universidad del País Vasco (UPV/EHU); así como por el proyecto Europeo FP7 Collaborative Research Project SECRET (SECURITY of Railways against Electromagnetic aTtacks), y el proyecto TEC2014-53785-R del Ministerio de Economía y Competitividad-fondos FEDER.

REFERENCIAS

- [1] V. Gungor, B. Lu, and G. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *Industrial Electronics, IEEE Transactions on*, vol. 57, no. 10, pp. 3557–3564, Oct 2010.
- [2] National Institute of Standards and Technology, "NISTIR 7628. Guidelines for Smart Grid Cyber Security," NIST, Technical Report, 2010. [Online]. Available: http://www.nist.gov/smartgrid/upload/nistir-7628_total.pdf
- [3] IEC TC57, *Communication networks and systems for power utility automation - Part 5: Communication requirements for functions and device models*, IEC 61850-5. Geneva, Switzerland, January 2013.
- [4] IEC 65C, *Industrial communication networks - High availability automation networks - Part 3: Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR)*, IEC 62439-3. Geneva, Switzerland, July 2012.
- [5] E. Haleplidis, K. Pentikousis, S. Denazis, J. H. Salim, D. Meyer, and O. Koufopavlou, "Software-Defined Networking (SDN): Layers and Architecture Terminology," RFC 7426 (Informational), Internet Engineering Task Force, Jan. 2015. [Online]. Available: <http://www.ietf.org/rfc/rfc7426.txt>
- [6] A. Willig, K. Mathews, and A. Wolisz, "Wireless technology in industrial networks," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1130–1151, June 2005.
- [7] P. Parikh, T. Sidhu, and A. Shami, "A comprehensive investigation of wireless lan for iec 61850 based smart distribution substation applications," *Industrial Informatics, IEEE Transactions on*, vol. 9, no. 3, pp. 1466–1476, Aug 2013.
- [8] P. Castello, P. Ferrari, A. Flammini, C. Muscas, P. Pegoraro, and S. Rinaldi, "Improving availability of distributed pmu in electrical substations using wireless redundant process bus," in *Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2014 IEEE International*, May 2014, pp. 239–244.

- [9] M. Rentschler and P. Laukemann, "Towards a reliable parallel redundant wlan black channel," in *Factory Communication Systems (WFCS), 2012 9th IEEE International Workshop on*, May 2012, pp. 255–264.
- [10] G. Cena, S. Scanzio, A. Valenzano, and C. Zunino, "An enhanced mac to increase reliability in redundant wi-fi networks," in *Factory Communication Systems (WFCS), 2014 10th IEEE Workshop on*, May 2014, pp. 1–10.
- [11] Open Networking Foundation (ONF), "OpenFlow Switch Specification, version 1.5.0," Open Networking Foundation (ONF), December 2014. [Online]. Available: <https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-switch-v1.5.0.noipr.pdf>
- [12] A. Lara, A. Kolasani, and B. Ramamurthy, "Network innovation using openflow: A survey," *Communications Surveys Tutorials, IEEE*, vol. 16, no. 1, pp. 493–512, First 2014.
- [13] E. Molina, E. Jacob, J. Matias, N. Moreira, and A. Astarloa, "Using Software Defined Networking to manage and control IEC 61850-based systems," *Computers Electrical Engineering*, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0045790614002626>
- [14] K.-K. Yap, T.-Y. Huang, M. Kobayashi, M. Chan, R. Sherwood, G. Parulkar, and N. McKeown, "Lossless handover with n-casting between wifi-wimax on openroads," *ACM Mobicom*, vol. 12, no. 3, pp. 40–52, 2009.
- [15] E. Molina, E. Jacob, J. Matias, N. Moreira, and A. Astarloa, "Availability Improvement of Layer 2 Seamless Networks Using OpenFlow," *The Scientific World Journal*, Jan 2015. [Online]. Available: <http://www.hindawi.com/journals/tswj/2015/283165/>
- [16] OpenDaylight, "Technical Overview." [Online]. Available: <http://www.opendaylight.org/project/technical-overview>
- [17] B. Pfaff, J. Pettit, T. Koponen, K. Amidon, M. Casado, and S. Shenker, "Extending networking into the virtualization layer," in *8th ACM Workshop on Hot Topics in Networks (HotNets)*, Oct. 2009. [Online]. Available: <http://conferences.sigcomm.org/hotnets/2009/papers/hotnets2009-final143.pdf>
- [18] M.-C. Chan, C. Chen, J.-X. Huang, T. Kuo, L.-H. Yen, and C.-C. Tseng, "Opennet: A simulator for software-defined wireless local area network," in *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, April 2014, pp. 3332–3336.
- [19] B. Lantz, B. Heller, and N. McKeown, "A network in a laptop: Rapid prototyping for software-defined networks," pp. 19:1–19:6, 2010. [Online]. Available: <http://doi.acm.org/10.1145/1868447.1868466>
- [20] P. Dely, A. Kassler, L. Chow, N. Bambos, N. Bayer, H. Einsiedler, C. Peylo, D. Mellado, and M. Sanchez, "A software-defined networking approach for handover management with real-time video in w lans," *Journal of Modern Transportation*, vol. 21, no. 1, pp. 58–65, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s40534-013-0007-x>

Uso de técnicas de emulación sobre la plataforma ns-3 para el análisis de QoS de servicios multimedia

Juan Arroyo, Luis Diez, Ramón Agüero
Departamento de Ingeniería de Comunicación,
Universidad de Cantabria
C.P. 39005

jaarsolares@gmail.com, ldiez@tlmat.unican.es, ramon@tlmat.unican.es

Resumen—En este trabajo se presenta un estudio acerca de la posibilidad de emplear técnicas de emulación para evaluar el comportamiento de servicios multimedia, en particular *streaming* de vídeo. Haciendo uso de la funcionalidad que incorpora la plataforma de simulación ns-3, así como de técnicas de virtualización, se describen los requerimientos, así como las principales limitaciones, de esta estrategia de análisis, de manera que proporcione una precisión adecuada. Finalmente, y para comprobar la viabilidad de la solución propuesta y sus posibilidades, se presenta un estudio de diferentes parámetros de QoS en aplicaciones de vídeo.

Palabras Clave—NS3, emulación, servicios de vídeo, LXC

I. INTRODUCCIÓN

La proliferación de terminales móviles de características avanzadas, así como el uso cada vez mayor por parte de los usuarios, está causando un gran cambio en los patrones de tráfico de redes móviles. De acuerdo con los últimos estudios [1], el tráfico generado por los servicios de *streaming* de vídeo está llamado a ser predominante en los próximos años. A raíz de esta demanda inesperada, aparece un importante reto para los operadores móviles. Para dar una respuesta apropiada al mismo, se está trabajando de manera intensa en la propuesta de diferentes algoritmos, técnicas y procedimientos en varios niveles.

Si algo tienen en común todas las soluciones propuestas es la necesidad de ser debidamente validadas, y de caracterizar las prestaciones que ofrecen. En este sentido, el uso de plataformas reales no es, en muchas ocasiones, una opción viable, ya sea porque la tecnología disponible no sea capaz de proporcionar la funcionalidad requerida o, simplemente, por la imposibilidad de recrear las condiciones que se asumen para la solución estudiada. Como consecuencia se ha popularizado el uso de herramientas de simulación, aunque en algunos casos son criticadas por su falta de realismo. Estas limitaciones se acentúan tanto en las capas más bajas (especialmente la capa física) como en el modelado de los servicios.

Este trabajo se centra en el estudio de técnicas empleadas por las herramientas de simulación capaces de dotar de mayor realismo al modelado de servicios. En concreto, el simulador de redes Network Simulator 3 (ns-3) [2] permite la conexión a nodos virtuales, capaces de ejecutar aplicaciones reales, externas al simulador, y que puedan a su vez conectarse a través de la tecnología de red subyacente proporcionada por el simulador. Esta funcionalidad extiende las capacidades típicas

de las plataformas de simulación más habituales, de modo que se puedan estudiar las interacciones entre servicios reales y una tecnología concreta (por ejemplo el efecto que un canal con errores tiene sobre un *streaming* de vídeo real). Además se mantiene la capacidad de reproducir de manera sistemática unas condiciones de comunicación concretas, lo que no sería posible sobre plataformas reales.

El resto de artículo se ha estructurado de la siguiente manera: en la Sección II se mencionan algunos trabajos que han utilizado la misma funcionalidad de simulación para analizar el rendimiento de servicios. A continuación, en la Sección III se presentan las conclusiones más relevantes que se han obtenido al estudiar el simulador ns-3 como herramienta de emulación; en primer lugar se han analizado los diferentes tipos de restricciones que presenta esta funcionalidad para, seguidamente, comprobar la viabilidad de ns-3 para analizar el comportamiento de servicios de vídeo reales. Finalmente, en la Sección IV se concluye el trabajo, adelantando una serie de líneas de investigación que se abordarán en el futuro y que se beneficiarán del conocimiento adquirido y de las posibilidades que proporciona la combinación de herramientas de simulación y técnicas de virtualización.

II. FUNDAMENTOS Y ANTECEDENTES

Las plataformas de simulación se han convertido en una herramienta clave para la comunidad investigadora. A pesar de la reciente aparición de varias plataformas de experimentación, las técnicas de simulación siguen ofreciendo un alto grado de flexibilidad, lo que las convierte en una alternativa muy atractiva para el estudio inicial de nuevas técnicas y conceptos, antes de su implementación sobre plataformas reales. Entre las ventajas de la simulación se pueden destacar la capacidad de analizar topologías complejas de una manera económica y efectiva o de reproducir de manera sistemática unas condiciones concretas, permitiendo establecer comparaciones de manera más apropiada. Como consecuencia, aún existe un gran número de trabajos de investigación que hacen uso de simuladores de red.

En todo caso, y a pesar de las ventajas mencionadas, la interacción con el *mundo real* es, en algunos casos, un aspecto crucial. Esto es particularmente adecuado cuando se analiza el impacto que una tecnología de red, entidad o solución concreta, puede tener sobre servicios específicos, por ejemplo aplicaciones con requisitos de tiempo real. A fin de

Tabla I: Interacciones del NS3 con el mundo real

Nodos	Redes		
	Real Simulado	Real Host NS3-EMU	Simulado NS3-TAP Simulación

satisfacer esta necesidad, el simulador ns-3 proporciona un conjunto de funcionalidades que permiten su interacción con entidades reales a diferentes niveles [3], tal como se muestra en la Tabla I. De esta manera, atendiendo a los diferentes requerimientos, se pueden conectar nodos simulados a una red real (NS3-EMU), nodos reales que se comunican con una infraestructura de red simulada (NS3-TAP), o recrear entornos totalmente simulados o reales. Este trabajo se ha centrado en la interacción de redes simuladas con nodos reales (indicado en la Tabla I como **NS3-TAP**) mediante el uso de técnicas de virtualización, lo que permite la ejecución de aplicaciones sobre infraestructuras de red simuladas. De aquí en adelante se usará el término emulación para referirse a esta aproximación.

Para poder interactuar con nodos reales, ns-3 proporciona una interfaz entre los entornos real y simulado, conocida como *NS3-TAP*, que puede conectarse a *bridges* virtuales instalados en la máquina nativa (*host*). La configuración más usada hace uso de Linux Containers (LXC) [4], ofreciendo la posibilidad de incluir nodos virtuales *ligeros*, en los que se pueden instanciar, por tanto, aplicaciones reales. Desde el punto de vista de la máquina nativa, el *bridge* virtual define una red dentro del *host*, a la que están conectados tanto el contenedor con la aplicación real, como el *NS3-TAP*. Además, el otro extremo del *NS3-TAP* se encuentra conectado a un terminal simulado, de modo que los datos enviados por el nodo virtual son redirigidos a aquel.

Por otro lado, la interacción con aplicaciones reales requiere que el tiempo de simulación (gestionado internamente por el simulador) esté alineado con el reloj del sistema. En la solución propuesta por ns-3 esto se consigue mediante un mecanismo de *scheduling* que se puede configurar para que haga uso del reloj del sistema (en lugar del que maneja internamente) a la hora de planificar los eventos que se producen y maneja en el transcurso de la simulación. A lo largo del documento se estudiará el rendimiento y las limitaciones presentadas por el simulador cuando se configura en tiempo real, prestando especial atención a la precisión de las medidas al utilizar una planificación de tiempo real.

En la literatura se pueden encontrar trabajos que han hecho uso de manera satisfactoria, sobre diferentes redes y con varios servicios, de la capacidad de emulación. Por ejemplo, en [5], [6] los autores utilizan la emulación proporcionada por ns-3 para estudiar el rendimiento de aplicaciones de vídeo sobre tecnología LTE, y analizan la influencia que los traspasos pueden tener sobre la Quality of Service (QoS) percibida por el usuario. Sin embargo, no analizan las posibles limitaciones del simulador para llevar a cabo dicho tipo de experimentos, ni detallan la configuración que se debería realizar.

Otros trabajos se han centrado en las aplicaciones. Por ejemplo, en [7] se desarrolló, sobre entornos emulados, una herramienta de medición de las QoS y Quality of Experience (QoE) de aplicaciones de vídeo, haciendo uso de una versión

modificada del conocido reproductor VLC. En este caso los autores sí mencionan las limitaciones del *scheduling* de tiempo real (en particular se centran en el parámetro *HardLimit*, que será estudiado posteriormente en la Sección III); sin embargo, los resultados que presentan son bastante limitados, y restringidos a escenarios concretos.

Finalmente, también existen otros trabajos que estudian de manera más concreta las limitaciones del simulador, aunque lo hacen desde una perspectiva diferente. Los autores de [8] proponen un sistema de emulación sobre ns-3 distribuido, que permita repartir la carga computacional entre un número de sistemas reales. Aunque esta solución permite reducir el tiempo de cómputo, existen topologías de red que no permitirían usar este enfoque. Un ejemplo relevante son las redes inalámbricas (incluyendo las celulares), en las que los diferentes nodos comparten el mismo medio físico. En estas condiciones, no sería posible poder mover nodos a otras máquinas, ya que no permitiría la correcta interacción en los niveles más bajos, especialmente la capa física.

III. EVALUACIÓN EXPERIMENTAL

A fin de evaluar la capacidad de la herramienta ns-3 para llevar a cabo emulación, en esta sección se presentan los resultados más destacados tras llevar a cabo una campaña de simulación con un doble objetivo; en primer lugar, en la Sección III-A se analizan, utilizando varias topologías con diferentes niveles de complejidad, las limitaciones presentadas por el simulador para ofrecer un nivel de precisión aceptable. A continuación, en la Sección III-B se presenta una evaluación de aplicaciones reales de vídeo en términos de QoS, estudiando el impacto que el uso de este tipo de servicios puede tener sobre el simulador. En ambos casos, se ha utilizado WiFi como tecnología de comunicaciones subyacente.

Cabe destacar que la máquina usada para obtener los resultados de este trabajo ha sido un ordenador ASUS S500CA, equipado con un procesador Intel Core i5 – 3317U@1.7GHz y 8 GB de memoria RAM. El sistema operativo usado ha sido Ubuntu 14.04LTS(64bits). Además, merece la pena destacar que todas las entidades que componen las topologías de red que se van a presentar (tanto entidades virtuales como simuladas) residen en el equipo descrito anteriormente.

III-A. Impacto del rendimiento de NS3 en el scheduling de tiempo real

En esta sección se presentan los principales resultados acerca de las limitaciones que presenta el simulador para utilizar emulación; en concreto, se ha estudiado la influencia del número de dispositivos físicos que pueden utilizarse manteniendo una precisión adecuada.

Para poder interpretar de manera más apropiada los resultados que se van a presentar posteriormente, es necesario explicar con más detalle el modo en que el simulador ns-3 se configura para utilizar técnicas de emulación. ns-3 es un simulador dirigido por eventos; así, las diferentes acciones que tienen lugar en el escenario que se esté analizando se *encolan*, según el instante de ejecución. En su configuración por defecto, ns-3 (y la mayoría de herramientas de esta índole) gestiona la ejecución de un evento tan pronto como haya finalizado el anterior, de manera que el tiempo de simulación es habitualmente sensiblemente menor al real. Por otro lado,

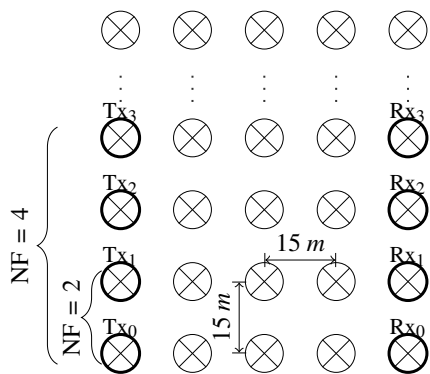


Figura 1: Ejemplo de escenario de simulación

para llevar a cabo simulación en tiempo real, ns-3 hace uso del reloj del sistema y, por tanto, los eventos se procesan de acuerdo a éste ¹. Así, si se produjera una gran cantidad de eventos en un intervalo reducido de tiempo, el simulador podría sobrecargarse, apareciendo un retardo entre el instante en que un determinado evento debió ser procesado y el momento en que realmente lo es; a este retardo se le llamará Simulation Delay (SD).

El simulador permite al usuario establecer un valor máximo del SD a través de la variable *HardLimit*; si el SD excede el valor establecido, la simulación se cancela. En esta sección se estudiará la evolución del SD para diferentes topologías de red. Se prestará especial atención al número de dispositivos físicos presentes en el escenario, ya que, al compartir el mismo canal físico, cada vez que se transmite o procesa un paquete, se genera un número elevado de eventos. En la práctica, ante el envío de un paquete de datos, todos los dispositivos conectados al mismo canal deberían comprobar si es necesario que lo procesen, lo que podrían provocar un retraso considerable del tiempo de ejecución. Es importante destacar que el efecto que se ha utilizado de ejemplo se produce en la capa física, por lo que no podría evitarse mediante el empleo de múltiples máquinas y ha sido, por otro lado, un gran obstáculo en los esfuerzos de paralelización llevados a cabo en el simulador.

Las topologías definidas en esta sección presentan una estructura de red regular, véase Figura 1. Los nodos han sido configurados de manera que empleen Ad hoc On-Demand Distance Vector Routing (AODV) como protocolo de encaminamiento, y la distancia entre las fuentes de tráfico y sus destinos correspondientes es de 4 saltos. Los flujos de tráfico utilizan una tasa constante y se han ejecutado simulaciones de 15 segundos de duración.

Como ejemplo ilustrativo, la Figura 2 muestra la evolución del valor del SD durante una simulación, para un número diferente de flujos y desplegando 50 nodos WiFi en la red. Los resultados revelan que, a pesar de la presencia de picos de retardo, el simulador es capaz de recuperarse, reduciendo la diferencia con el tiempo real. Este comportamiento es consecuencia de la acumulación de eventos en algunos momentos (por ejemplo, cuando los nodos envían los mensajes

¹Aunque el orden de los eventos no se ve alterado, en la configuración en tiempo real pueden encontrarse períodos de espera, hasta que el siguiente evento deba ser procesado.

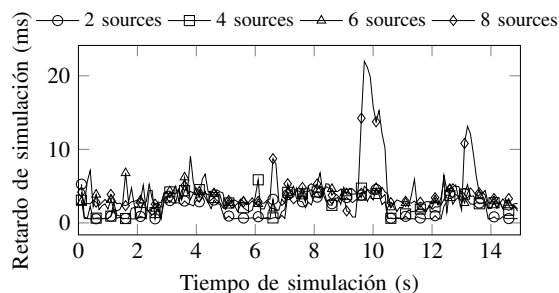


Figura 2: Evolución del retardo de simulación en un único experimento con 50 nodos

de HELLO utilizados por el protocolo de encaminamiento). Por tanto, se puede ver que si se fijara un valor estricto de retardo (por medio del parámetro *HardLimit*) se descartarían algunos escenarios, incluso si el valor se superara únicamente en un momento puntual de la simulación.

La Figura 3 muestra el comportamiento estadístico del simulador se presenta, representando la Cumulative Distribution Function (CDF) del SD cuando se utilizan un número diferente de flujos de tráfico en el escenario. Los resultados se han obtenido midiendo el valor de SD cada 100 ms durante 20 simulaciones, dando lugar a 3000 muestras, a partir de las cuales se han obtenido las funciones de probabilidad acumulada, para cada una de las configuraciones que se analizaron.

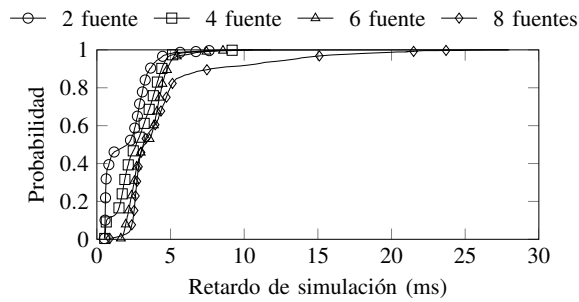
Como se puede observar, ns-3 es capaz de llevar a cabo la simulación con una precisión adecuada de redes ad-hoc WiFi en las que se despliegan hasta 50 nodos. En este sentido, la Figura 3a muestra que más del 90% de las muestras corresponden a valores de SD inferiores a 5ms, excepto cuando el número de flujos de tráfico crece hasta 8. En este caso, el tráfico presente en la red afecta el precisión de la funcionalidad de tiempo real.

Por otro lado, el incremento del número de dispositivos físicos, véase Figura 3b, tiene un alto impacto en el valor de SD. Como se pone de manifiesto, el 90% de las muestras con 2, 4 y 6 fuentes tienen un retardo inferior a los 500ms (100 veces superior a los tiempos medidos para el escenario anterior), mientras que los valores en el caso en que se establecen 8 flujos de tráfico no serían válidos en la mayoría de los casos.

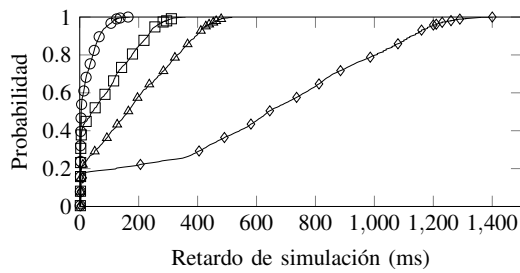
A la luz de los resultados observados, se pueden extraer algunas conclusiones interesantes. La primera es que la configuración de un valor máximo de SD de forma estricta puede no ser siempre la opción más conveniente, ya que la evolución temporal de este parámetro fluctúa de manera muy apreciable a lo largo del tiempo, siendo capaz de recuperarse tras incrementos puntuales del retardo. La segunda conclusión, muy relevante en términos del dimensionado del escenario, indica que el número de dispositivos físicos es un aspecto clave si se pretende llevar a cabo experimentos emulados.

III-B. Caracterización de una sesión de vídeo

En esta sección se presenta la evaluación del comportamiento de una aplicación de vídeo real, haciendo uso de la funcionalidad de emulación de la plataforma ns-3. Para ello se han creado instancias del reproductor de vídeo VLC en



(a) 50 dispositivos físicos



(b) 100 dispositivos físicos

Figura 3: Retardo de simulación con diferente número de fuentes de tráfico

los dispositivos virtuales LXC y se han generado sesiones con una tasa de muestreo y de generación de imágenes de 44100 Hz y 20 fps respectivamente. Con el fin de tener un mayor control sobre la red, las coberturas de los nodos WiFi se han fijado de manera que se eviten las colisiones entre paquetes pertenecientes a diferentes flujos de tráfico; en concreto, se ha hecho uso de un modelo de propagación (disco) de radio constante de 15 m . Cabe destacar que, aunque se eviten colisiones, la implementación del protocolo 802.11 por parte del simulador no evita que los diferentes dispositivos contienda por el canal, incluso si se encuentran fuera del radio de cobertura.

En primer lugar se ha llevado a cabo un conjunto de simulaciones con el objetivo de estudiar los efectos que el tráfico real pueda tener sobre el rendimiento en tiempo real del simulador. Para ello ha establecido un sencillo escenario en el que el número de flujos de tráfico de vídeo se ha ido incrementando paulatinamente y en el que no existen errores en el canal; la Figura 4 muestra un ejemplo ilustrativo con 4 flujos. Como se puede observar, cada flujo de vídeo requiere una pareja de dispositivos virtuales, y que los nodos, dentro del simulador, que forman parte del intercambio de datagramas pertenecientes a diferentes flujos de vídeo no se interfieren entre sí.

Todas las métricas para analizar el comportamiento (la QoS) sobre este escenario, a excepción del SD, que considera la topología de red completa, se han medido en la comunicación entre la primera pareja de dispositivos (1 y 2 en la Figura 4), de manera que se comprueba la influencia de introducir una carga mayor de tráfico. Además, se han llevado a cabo 20 experimentos independientes de cada una de las configuraciones, a fin de asegurar la validez estadística de los

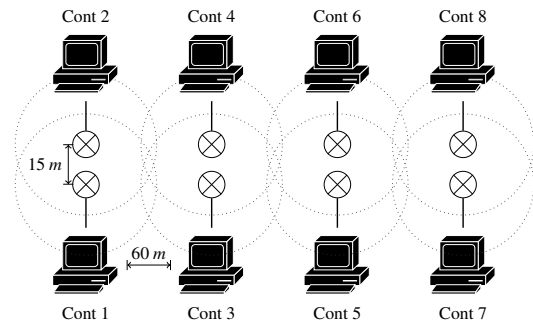


Figura 4: Escenario genérico con 4 flujos de vídeo. Los ordenadores representan LXC que se conectan a nodos WiFi en el entorno simulado

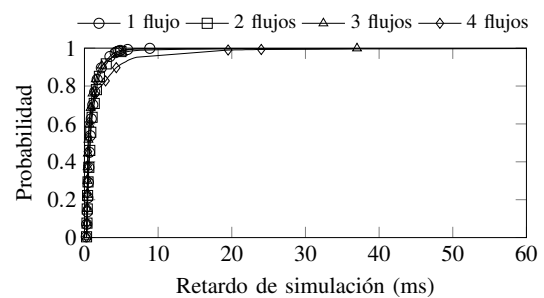


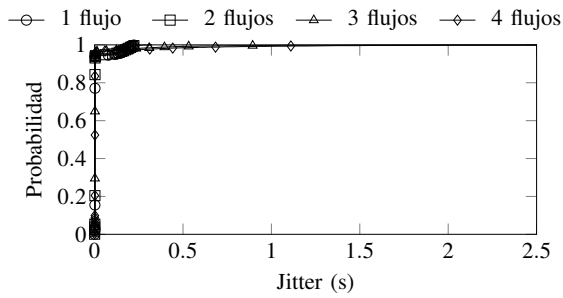
Figura 5: Retardo de simulación (SD) ante tráfico de vídeo real

resultados.

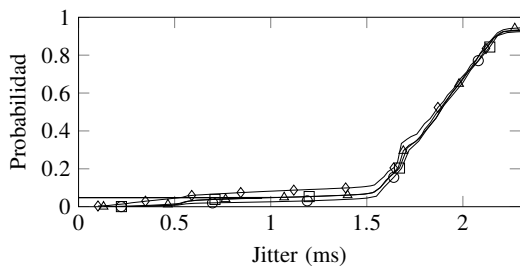
En primer lugar, la Figura 5 muestra la CDF del parámetro SD, para diferentes situaciones, en las que se va incrementando el número de flujos de tráfico activos. Al igual que en la Sección III-A, el valor de SD se ha obtenido con una frecuencia de 100 ms en cada simulación. Como puede observarse, se consigue ejecutar la emulación con valores de SD inferiores a 30 ms hasta con 4 fuentes de tráfico. Al comparar estos resultados con los obtenidos al incrementar el número de dispositivos desplegados en la red, queda claro que este último es el factor que en mayor medida podría limitar el uso de este tipo de análisis.

Una vez analizado el rendimiento al utilizar aplicaciones reales, en las que el tráfico se corresponde con el que se tendría en un escenario real, se analizan seguidamente un conjunto de métricas ilustrativas del QoS de los servicios de vídeo sobre redes multi-salto, utilizando WiFi como tecnología subyacente. La Figura 6 muestra el *jitter* de los paquetes generados por la aplicación de vídeo, medido como el tiempo transcurrido entre la recepción de dos paquetes consecutivos. Los resultados ponen de manifiesto la escasa influencia del número de flujos de tráfico presentes en el escenario sobre este parámetro, ya que la mayoría de las muestras observadas tienen un valor cercano al 0; así, como puede observarse con mayor claridad en la Figura 6b, más del 90% de los retardos son inferiores a 2.5 ms . A la hora de procesar las estadísticas recogidas se observa que nuevamente se vuelven a producir, de manera muy puntual, valores de retardo ‘elevados’, incluso cercanos a 1 segundo.

También se ha analizado el retardo extremo a extremo que



(a)



(b)

Figura 6: Función de probabilidad acumulada del *jitter* con diferente número de flujos

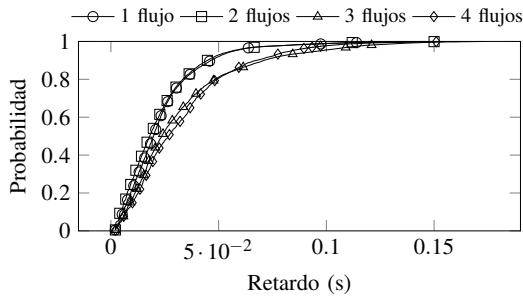
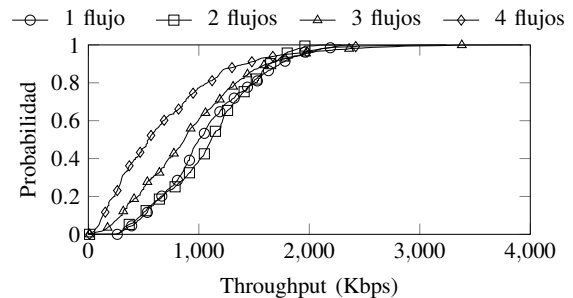


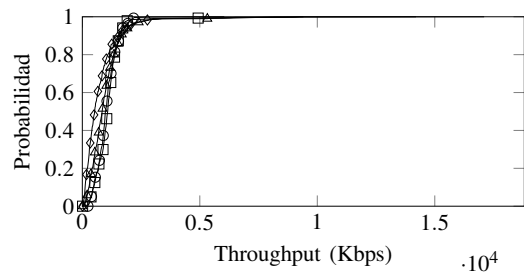
Figura 7: Función de probabilidad acumulada del retardo con diferente número de flujos

presenta la red, reflejado en la Figura 7. En este caso los resultados ponen de manifiesto un incremento notable del retardo a medida que se introducen más flujos de tráfico. Como se comentó anteriormente, a pesar de que el modelo de propagación evita las colisiones, no evita la contienda por el canal, lo que se evidencia al utilizar 3 y 4 servicios de vídeo, ya que el mayor número de paquetes generados tiene como consecuencia un aumento significativo de los eventos de contienda que el simulador tiene que manejar.

Como última métrica a caracterizar en este primer escenario se ha obtenido el caudal visto por la aplicación, tanto en transmisión como en recepción, de nuevo únicamente para los dos nodos correspondientes al primer flujo de tráfico. La Figura 8 muestra la CDF correspondiente al *throughput* de dicho servicio de vídeo. Los resultados permiten establecer la influencia de una mayor presencia de flujos de tráfico (lo que deriva en un mayor número de instancias del reproductor VLC), con un impacto en el rendimiento observado tanto en transmisión como en recepción. Sin embargo, se puede intuir



(a) Tasa de transmisión



(b) Tasa de recepción

Figura 8: Función de probabilidad acumulada del *throughput* con diferente número de flujos

que este hecho sea consecuencia de la capacidad de cómputo del sistema utilizado, ya que las tasas de tráfico se estiman en el punto de entrada al simulador, por lo que en principio el ns-3 no debería afectarlas de manera directa.

Para concluir con el análisis, y con el fin de ilustrar una de las principales ventajas de esta metodología, se ha llevado a cabo una evaluación del comportamiento del servicio de *streaming* de vídeo sobre un escenario en el que aparezcan errores en el canal. En este caso, el escenario considera un servidor de vídeo que envía simultáneamente (*broadcasting*) un mismo vídeo a dos clientes, con los que tiene unas condiciones de comunicación diferentes (mayor y menor calidad en los caminos correspondientes). La Figura 9 ilustra el escenario que se ha desplegado; los receptores **Cliente1** y **Cliente2** se encuentran a 4 y 2 saltos de distancia con el servidor, respectivamente; además, las condiciones de los canales, establecidas por la la tasa de error de trama (*Frame Error Rate*, FER) son peores en la ruta que conecta el servidor con el **Cliente1**. En este caso se comentan los resultados observados tras la realización de una simulación con una duración de varios minutos, lo que permite estudiar el comportamiento en situaciones más realistas.

En la Figura 10 se recoge el caudal observado, tanto en transmisión como en recepción (para los dos clientes). Se puede observar que el *throughput* en recepción en el **Cliente2** es ligeramente superior al correspondiente al **Cliente1**, debido a las peores condiciones en la ruta que une éste con el servidor (con un mayor número de saltos, en los que la FER también es superior). Por otro lado, también queda patente que, mientras el rendimiento medido en los clientes es bastante estable (alrededor de 1 Mbps), el *throughput* en el transmisor (servidor) oscila entre 1 y 5 Mbps. Es importante destacar que este tipo de comportamiento, así como los paquetes que la aplicación en los clientes envía al servidor, no podrían ser

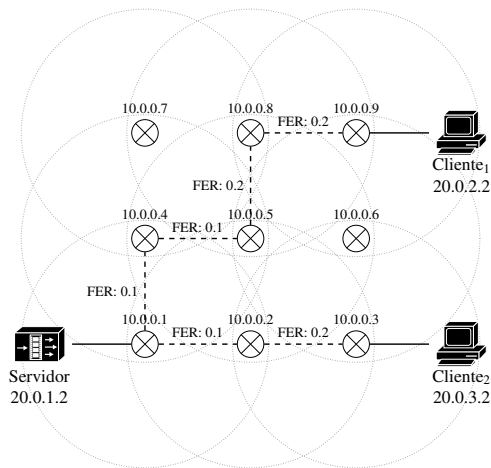


Figura 9: Escenario con errores en el canal

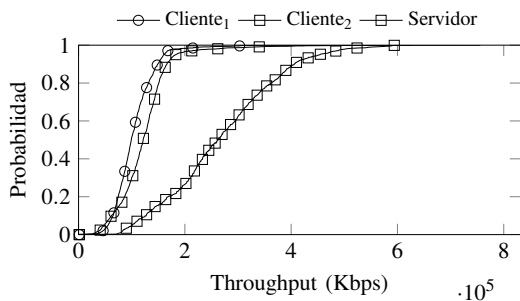


Figura 10: Función de probabilidad acumulada del *throughput* en el escenario con errores

correctamente reflejados a partir de los modelos sintéticos de tráfico que habitualmente emplean los simuladores, lo que pone de manifiesto que la metodología propuesta sí que consigue otorgar un mayor nivel de realismo en los servicios utilizados para llevar a cabo los análisis.

Como complemento al rendimiento, la Figura 11 muestra el porcentaje de paquetes recibidos por cada uno de los clientes, así como el porcentaje de pérdidas totales en el escenario, es decir, los paquetes que habiendo sido enviados por el servidor no son recibidos por ningún cliente. Como se puede observar, se produce una pérdida significativa de paquetes, en el conjunto del escenario, próxima al 20%. Por otro lado se puede ver que el porcentaje de paquetes recibidos por cada cliente no es igual, ya que, como consecuencia de las peores condiciones de la ruta que le une al servidor, el **Cliente1** recibe una cantidad menor de paquetes.

Gracias a la funcionalidad que aporta la metodología presentada en este trabajo, es posible analizar, de manera *subjetiva*, la influencia de las condiciones de la red sobre la percepción que tendría un usuario al visualizar el vídeo. Así, la Figura 12 muestra sendas capturas de pantalla obtenidas en las aplicaciones de ambos clientes. De manera genérica, se puede decir que, con valores de FER moderados (0.1 y 0.2), se pone de manifiesto una degradación paulatina de ambos vídeos, consecuencia tanto de las condiciones hostiles de los diferentes canales y del hecho de que se traten de un canal compartido. En las capturas que se han utilizado se ha tratado de plasmar la tendencia general, observándose

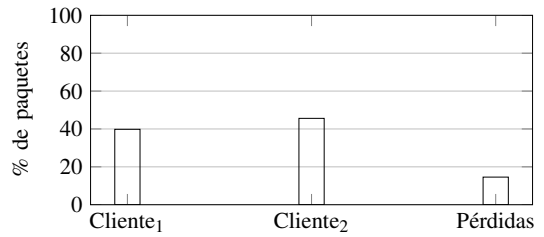
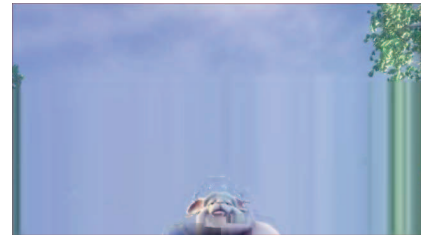


Figura 11: Porcentaje de paquetes recibidos por cada cliente y el total de pérdidas



(a) Cliente 1



(b) Cliente 2

Figura 12: Capturas de vídeo representativas de ambos clientes

que la degradación sufrida por el **Cliente1** es más acusada, poniéndose de manifiesto el efecto negativo que tienen las comunicaciones sobre redes inalámbricas multi-salto.

IV. CONCLUSION

En este trabajo se ha propuesto una metodología que plantea el uso de técnicas de emulación para el análisis de servicios de vídeo sobre plataformas de simulación. Para ello se ha hecho uso de las funcionalidades que ofrece la herramienta ns-3, que permiten su interacción con sistemas reales. En concreto se ha llevado a cabo un estudio ilustrativo del comportamiento de servicios de *streaming* de vídeo reales al ser utilizadas sobre redes inalámbricas multi-salto, utilizando tecnología WiFi.

En primer lugar se han analizado las limitaciones que presenta el simulador para acometer de manera apropiada las tareas de emulación en tiempo real, así como otros parámetros que influyen en la precisión de los resultados que se pueden obtener. Este primer análisis permite concluir que el número de dispositivos físicos que se despliegan en el escenario tiene un gran impacto sobre la capacidad del simulador ns-3 para llevar a cabo análisis en tiempo real. Además se ha podido observar que su influencia es mayor que la de un incremento en la carga de tráfico total en el sistema, debido a que incluso los nodos que no participan activamente en las simulaciones generan una serie de eventos que el simulador debe atender.

Por otro lado, también se han observado comportamientos puntuales que bien podrían ser debidos a las limitaciones del equipo sobre el que se desarrollaron los experimentos. Para solventar esta limitación, una posibilidad sería que, en caso de que sea necesario instanciar un número elevado de aplicaciones, éstas podrían trasladarse a otras máquina; así, se aliviaría la carga computacional del equipo en el que se simula la red inalámbrica.

Finalmente se ha realizado un análisis de un servicio real de vídeo (utilizando VLC), tanto en condiciones ideales como introduciendo errores en el canal. Los resultados obtenidos muestran el comportamiento que podría haberse anticipado, lo que confirma la capacidad de la metodología propuesta para llevar a cabo la evaluación de aplicaciones reales utilizando despliegues de red ‘simulados’. Por otro lado se han observado asimismo comportamientos de los patrones de tráfico generados por las aplicaciones inesperados, pero que otorgan un mayor nivel de realismo a la integración llevada a cabo, ya que no podrían ser reproducidos en ningún caso por entornos completamente simulados. Por otro lado, la capacidad de utilizar de manera sistemática un escenario con condiciones de la comunicación reproducibles (saltos, distancias, probabilidad de error, etc.) otorga al método propuesto de una gran flexibilidad, siendo este uno de los objetivos que se buscaban con la metodología propuesta.

En lo que se refiere a la investigación que se acometerá en el futuro, se pretende que este trabajo sirva de apoyo a la hora de analizar el comportamiento de diferentes soluciones sobre redes inalámbricas multi-salto (p.e. algoritmos de encaminamiento, técnicas de *network coding*), enriqueciéndose con la posibilidad de utilizar aplicaciones reales. Por otro lado, como extensión al trabajo presentado, se estudiará la capacidad de emulación del simulador ns-3 utilizando otras tecnologías de comunicación, en concreto LTE, dada la relevancia que va a tener en los próximos años.

AGRADECIMIENTOS

Los autores querrían expresar su agradecimiento al Gobierno de España por su financiación en el proyecto “*Connectivity as a Service: Access for the Internet of the Future*”, COSAIF (TEC2012-38574-C02-02)

REFERENCIAS

- [1] CISCO, “Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update is part of the comprehensive Cisco VNI Forecast, an ongoing initiative to track and forecast the impact of visual networking applications on global networks;” February 2014.
- [2] Network simulator 3. [Online]. Available: <http://www.nsnam.org/>
- [3] How to make ns-3 interact with the real world. [Online]. Available: https://www.nsnam.org/wiki/HOWTO_make_ns-3_interact_with_the_real_world
- [4] How to use linux containers to set up virtual networks. [Online]. Available: https://www.nsnam.org/wiki/HOWTO_Use_Linux_Containers_to_set_up_virtual_networks
- [5] T. Molloy, Z. Yuan, and G.-M. Muntean, “Real time emulation of an lte network using ns-3,” in *Irish Signals Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CICT 2014)*. 25th IET, June 2014, pp. 251–257.
- [6] A. Fouda, A. N. Ragab, A. Esswie, M. A. Marzban, A. Naser, M. Rehan, and A. S. Ibrahim, “Real time video streaming over ns3 based emulated lte networks,” in *International Journal of Electronics Communication and Computer Technology (IJECCCT) Volume 4 Issue 3*, May 2014.

- [7] J. Bustos-Jimenez, R. Alonso, C. Faundez, and H. Meric, “Boxing experience: Measuring qos and qoe of multimedia streaming using ns3, lxc and vlc;” in *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, Sept 2014, pp. 658–662.
- [8] A. Alvarez, R. Orea, S. Cabrero, X. G. Pañeda, R. García, and D. Melendi, “Limitations of network emulation with single-machine and distributed ns-3,” in *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*, ser. SIMUTools '10. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2010, pp. 67:1–67:9. [Online]. Available: <http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8630>

Uso de técnicas de negociación automática para la asignación de canales en IEEE 802.11

Enrique de la Hoz¹, Jose Manuel Gimenez-Guzman¹, Ivan Marsa-Maestre¹ y David Orden Martín²

¹ Departamento de Automática

² Departamento de Física y Matemáticas

Universidad de Alcalá

Edificio Politécnico, Campus Universitario, 28805 Alcalá de Henares, Madrid.

{enrique.delahoz, josem.gimenez, ivan.marsa, david.orden}@uah.es

Resumen—Las redes inalámbricas actuales deben dar soporte a cada vez más servicios telemáticos que demandan mayores prestaciones. Puesto que el espectro radioeléctrico en el que estas redes se basan es limitado, aumentar su eficiencia es ya una exigencia. En este trabajo se aborda el problema de la asignación de canales a puntos de acceso IEEE 802.11n con el objetivo de que los clientes de la red vean mejorada su experiencia de utilización. Las técnicas que se emplean con tal objetivo se basan en la negociación automática, comparándose con otras alternativas. Los resultados muestran que las técnicas de negociación son muy adecuadas para resolver el problema en cuestión, siendo capaces de obtener prestaciones superiores al resto de métodos con los que se han comparado.

Palabras Clave—Redes inalámbricas, IEEE 802.11n, negociación automática, asignación de canales.

I. INTRODUCCIÓN

En el contexto de las telecomunicaciones, el recurso más preciado por su escasez actual es el espectro radioeléctrico. Muchos son los motivos para esta escasez. En primer lugar, cada vez son más las tecnologías que requieren usar el espectro radioeléctrico y, además, en los mismos lugares. Por otro lado, el número de usuarios que hace uso de las mismas está sufriendo un vertiginoso aumento en la última década. Además, los usuarios cada vez requieren un mayor ancho de banda de las redes inalámbricas que emplean, de manera que cada vez es más difícil que las redes de comunicaciones inalámbricas sean capaces de ofrecer servicios telemáticos solventemente y con ciertas garantías de servicio. Por lo tanto, se puede concluir que es una necesidad real el hacer un uso más eficiente del espectro radioeléctrico. Muchas son las soluciones que se están dando al problema de la escasez del espectro radioeléctrico a todos los niveles. Entre estas soluciones encontramos algunas basadas en mejorar la eficiencia espectral [1] y los algoritmos de detección y corrección de errores [2] o en diseñar nuevas antenas que permitan aprovechar la propagación multicamino para recibir señales de una manera más eficiente (MIMO) [3]. Otras se centran en, para una determinada tecnología, encontrar mecanismos de coordinación para los dispositivos que conviven en la misma banda de frecuencia, de forma que se minimicen las interferencias y se aumente el rendimiento. Es lo que se conoce como problema de la asignación de frecuencias, o *frequency assignment problem (FAP)* [4]. La asignación de frecuencias se ha abordado de forma centralizada desde la perspectiva de optimización [5], y desde la ingeniería con heurísticas distribuidas. Entre esas heurísticas, probablemente una de las más ampliamente desplegada sea LCCS (*Least*

Congested Channel Search) [6], que es la que se utiliza en las redes IEEE 802.11. De hecho, la asignación de canales en IEEE 802.11 constituye un subconjunto interesante de los problemas FAP, dada la ubicuidad de este tipo de redes y el crecimiento exponencial del número de dispositivos que las emplean en los últimos años.

En este artículo se pretende abordar el problema de la asignación de frecuencias (canales) en redes IEEE 802.11 desde la perspectiva de las técnicas de negociación automática. La negociación automática ha demostrado ser eficaz para el apoyo a la toma de decisiones en escenarios donde es necesario llegar a acuerdos de forma rápida y conciliando intereses en conflicto [7]. Las potenciales aplicaciones de la negociación automática van desde el comercio electrónico [8] a la resolución de problemas complejos de distribución de tareas, compartición de recursos o diseño cooperativo [9], [10], [11], [12]. Una de las ventajas de la negociación automática es que tiene en cuenta el conflicto de intereses desde el principio, lo que permite llegar a soluciones (acuerdos) más estables y menos proclives a que los agentes implicados (por ejemplo, los usuarios de los puntos de acceso IEEE 802.11 en un determinado edificio) decidan desviarse de la solución socialmente óptima en favor de otra que maximice su utilidad individual.

En particular, este artículo pretende contrastar la hipótesis de que las técnicas de negociación no lineal, en las que los autores han trabajado de forma extensiva [13], [14], pueden ser aplicadas al escenario concreto de la asignación de canales en IEEE 802.11. Este trabajo contribuye a este objetivo en las siguientes formas:

- Presentamos un modelado abstracto del problema inspirado en el problema matemático del coloreado de grafos (Sección II-A).
- Planteamos un modelo de utilidad de las asignaciones de frecuencia que tiene en cuenta la influencia de la interferencia entre los diferentes canales IEEE 802.11 en la relación señal a ruido de un receptor en un momento dado (Sección II-D).
- Realizamos una primera aproximación al problema de asignación de frecuencias en IEEE 802.11 desde la perspectiva de la negociación, en un escenario con dos proveedores de servicio (Sección III).

Para validar la hipótesis de partida y valorar el efecto de nuestras contribuciones, se ha realizado un conjunto de experimentos, que se describen y analizan en la Sección IV.

Finalmente, el artículo concluye con una discusión de los resultados obtenidos y el planteamiento de líneas futuras de trabajo.

II. MODELADO DEL PROBLEMA

A. Modelado abstracto basado en coloreado de grafos

Los grafos se encuentran entre las herramientas más utilizadas para modelar el problema de asignación de frecuencias, debido a la relación de este problema con el de coloreado de grafos, ampliamente estudiado en la comunidad matemática [15]. En el coloreado de grafos se considera un grafo abstracto, dado como un conjunto de vértices junto con ciertas aristas entre ellos, y se tiene por objetivo asignar un color a cada vértice de modo que se utilice el menor número posible de colores evitando las aristas monocromáticas.

En el modelo utilizado habitualmente, los nodos del grafo representan elementos a los que se debe asignar una frecuencia y las aristas representan aquellos pares de elementos que no deben recibir la misma frecuencia. De este modo, los colores juegan el papel de las frecuencias y las aristas heterocromáticas aseguran pares de elementos con frecuencias distintas.

Pese a lo extendido de este modelo, Tragos et al. [16] concluyen que resulta insuficiente, puesto que no refleja toda la información. Por ejemplo, sugieren que se debería incorporar al grafo la información correspondiente a interferencias entre canales adyacentes. En esta línea, el presente trabajo propone la utilización de un grafo que modela con más fidelidad el problema.

Como elementos del grafo, distinguimos dos tipos de vértices y dos tipos de aristas. En cuanto a los vértices, consideramos puntos de acceso (AP) y clientes (C). En cuanto a las aristas, por un lado cada cliente está unido a su AP más cercano (con el fin de reflejar que dicho cliente utilizará el canal que se asigne a su AP más cercano) y por otro lado uniremos también algunos pares de nodos cuando la distancia entre ellos sea menor que el correspondiente radio de interferencia R (para reflejar las interferencias): los pares de tipo AP-AP se unirán siempre que se cumpla esa condición de distancia, los pares de tipo AP-C solo cuando el cliente no esté asociado a ese AP, y los pares de tipo C-C solo cuando ambos clientes estén asociados a distintos APs.

Además, con el fin de modelar la potencia de la interferencia entre dos elementos, asignamos un peso a cada arista basándonos en tres factores. En primer lugar, consideramos un peso para cada par de colores ij , que puede interpretarse como la interferencia entre el color i y el color j . Es interesante destacar que el problema usual de coloreado solo considera el caso particular de interferencias entre pares de vértices con el mismo color, mientras que nuestra extensión del problema permite considerar también interferencias entre colores adyacentes o en un cierto rango de distancias.

En segundo lugar modulamos los pesos anteriores, considerando la distancia entre los extremos de las aristas. De este modo, el peso de una arista coloreada ij será diferente dependiendo de cómo de cercanos estén sus extremos. Esto supone una nueva extensión del problema usual de coloreado, puesto que ahora los vértices tienen unas posiciones concretas y eso hace que nuestro grafo pase a ser geométrico en lugar de abstracto.

En tercer lugar introducimos en los pesos el efecto del tiempo, considerando que los vértices alternan intervalos en los que se encuentran activos con otros intervalos de tiempo en los que están inactivos y por tanto no afectan al coloreado. En la siguiente sección se describe cómo se tienen en cuenta los efectos de los tres factores citados sobre el escenario de redes IEEE 802.11n.

B. Efectos de la propagación en la cobertura

Para poder evaluar un determinado coloreado de un escenario IEEE 802.11 se debe incorporar y modelar el efecto que tienen sobre las señales tanto la propagación como las interferencias. Para el modelo de propagación, en [17] se define que las pérdidas de potencia de la señal de radiofrecuencia, expresadas en dB y considerando que las antenas están próximas al suelo (entre 1 y 2.5m), pueden calcularse mediante la siguiente ecuación:

$$P_{loss} = 40\log_{10}d + 20\log_{10}f - 20\log_{10}(h_t h_r), \quad (1)$$

donde d es la distancia expresada en metros, f es la frecuencia de la señal expresada en GHz y h_t (h_r) es la altura de la antena de transmisión (recepción), también expresada en metros. La Ec. 1 puede simplificarse para su uso en la banda de frecuencias de 2.4 GHz, empleada por IEEE 802.11b/g/n, obteniéndose:

$$P_{loss} = 7.6 + 40\log_{10}d - 20\log_{10}(h_t h_r). \quad (2)$$

Un caso particular especialmente interesante del modelo de propagación empleado es el relativo al caso límite que supone recibir una señal con una potencia igual a la sensibilidad del receptor, es decir:

$$P_t + G_t + G_r - L - P_{loss} = S, \quad (3)$$

donde P_t representa la potencia de transmisión (en dBm), G_t (G_r) representa la ganancia de la antena de transmisión (recepción) (en dB), L representa las pérdidas debidas a paredes, ventanas y obstáculos en la propagación (en dB), P_{loss} representa las pérdidas por la propagación (en dB) y S es la sensibilidad del receptor (en dBm). Si empleamos la Ec. 2 podremos conocer la distancia a partir de la cual las señales no pueden ser percibidas por el receptor. Dicha distancia (expresada en metros) puede calcularse como:

$$R = 10^{\frac{P_t + G_t + G_r - L - S - 7.6 + 20\log_{10}(h_t h_r)}{40}}. \quad (4)$$

El interés del cálculo de R según la Ec. 4 reside en que dicha distancia constituirá la distancia máxima a la que puedan estar dos nodos para que se conecten en el grafo de interferencias.

C. Efectos de las interferencias

Para ser capaces de cuantificar la calidad de la señal recibida se deben considerar las señales interferentes (interferencias), entendidas estas como las señales no deseadas que reciben los diferentes elementos de la red y que dificultan la correcta recepción de la señal deseada. Así, para el problema en cuestión, es necesario cuantificar la potencia de las señales interferentes en todos los puntos de recepción de señal, es decir, tanto en los terminales como en los APs. En primer lugar, para modelar dichas interferencias hay que recordar que se

ha asumido que solo existen interferencias de otros emisores IEEE 802.11, no considerando otras posibles interferencias de otras tecnologías que operen en la banda de frecuencias de 2.4GHz. No sería complejo añadir dichas interferencias al modelo que se va a describir a continuación, pero no se han incluido porque su aparición es ajena al problema que se trata en este artículo. Por otro lado, puesto que la distancia es un parámetro básico en la propagación también de señales interferentes, se han despreciado las interferencias que provienen de fuentes que se encuentran lejos del punto de recepción, de manera que la sensibilidad del receptor no alcance a recibirlas, siguiendo el grafo de interferencia que se ha descrito en la Sección II-A. Finalmente, se ha asumido que los terminales conectados a un mismo AP no interfieren entre sí, puesto que dichas comunicaciones están coordinadas por el AP.

En cada nodo, la calidad de la señal recibida depende de todas las interferencias recibidas, las cuales han sido transmitidas por otros APs y terminales. Además de la distancia (que influye en las pérdidas de la señal, Ec. 2), las interferencias también resultan afectadas por otros dos factores: el índice de actividad (Ψ) y el índice cocanal (δ). El índice de actividad representa la fracción del tiempo en que una señal interferente se está recibiendo respecto al tiempo total. Dicho índice modela el hecho de que resultará más perjudicial una señal interferente que se emita el 60% del tiempo que otra que se emite el 10% del tiempo. En concreto, se han considerado dos índices de actividad diferentes: para los APs (Ψ_{AP}) y para los terminales (Ψ_T). Puesto que el tráfico de Internet es mayoritariamente asimétrico en sentido *downstream*, se ha considerado que $\Psi_{AP} > \Psi_T$, en concreto se ha usado $\Psi_{AP} = 0,5$ y $\Psi_T = 0,2$. Esta elección se ha hecho para considerar la asimetría del tráfico, si bien sus valores absolutos dependerán de factores muy variables como la aplicación o la hora del día. Por otro lado, el índice cocanal (δ) representa el solape existente entre las diferentes bandas de frecuencia empleadas por la tecnología IEEE 802.11 cuando esta opera en la banda de 2.4GHz. Para la inclusión de este efecto, se han empleado los valores medidos para este índice para IEEE 802.11n en [18], donde se muestra una matriz en la que cada valor (i, j) representa la interferencia en el canal i que produce una transmisión en el canal j . Sintetizando lo expuesto, si un nodo que opera en el canal i recibe la interferencia de un nodo que opera en el canal j , la potencia de dicha interferencia viene dada por:

$$I = P_t + G_t + G_r - L - P_{loss} + \delta(i, j) + \Psi, \quad (5)$$

donde todos los valores se expresan en dB, incluido Ψ , que será Ψ_{AP} o Ψ_T si la señal interferente proviene de un AP o terminal cliente, respectivamente. Una vez conocido el modelado de las señales interferentes, se puede determinar la relación señal a ruido en cada terminal ($SINR_i$, para el terminal i -ésimo) como el cociente entre la señal recibida y la suma de las interferencias recibidas. Nótese que cada AP tendrá un valor de $SINR$ para cada uno de los terminales con los que esté conectado. En este caso, se ha considerado que su $SINR$ será el menor de todos ellos, es decir, el peor de los casos.

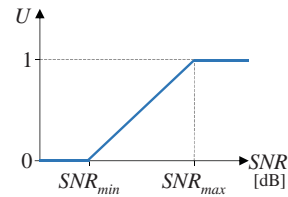


Fig. 1. Relación entre la utilidad y $SINR$.

D. Utilidad de las soluciones

Para cuantificar la bondad de los diferentes coloreados de la red realizados, se ha empleado el concepto de utilidad, el cual está muy directamente relacionado con la $SINR$ y el régimen binario alcanzado [19].

Así, la utilidad que percibe el nodo i -ésimo (U_i) es un valor entre 0 y 1, siendo 0 el caso en el que la recepción es de baja calidad y el terminal (AP) no se mantiene constantemente conectado con su AP (terminal). Por otro lado, el caso en el que $U_i = 1$ se dará cuando la calidad de la señal es excelente. La relación entre la utilidad y la $SINR$ se puede observar en la Fig. 1.

Para definir los valores umbrales de $SINR$ que definen el máximo y el mínimo de la utilidad, se han empleado los valores medidos en [20], siendo estos $SINR_{min} = 10dB$ y $SINR_{max} = 40dB$.

Finalmente, la utilidad de un coloreado específico de la red se calcula como la suma de las utilidades de todos los nodos que componen dicha red, es decir:

$$U = \sum_{\forall i} U_i. \quad (6)$$

III. TÉCNICAS DE NEGOCIACIÓN AUTOMÁTICA PARA LA SELECCIÓN DE CANALES

Tal y como apuntábamos en la introducción, en este artículo nos planteamos abordar el problema de la asignación de frecuencias en IEEE 802.11 desde la perspectiva de la negociación automática. Aunque la negociación automática es un campo de gran diversidad, para el contexto de este artículo podemos considerar un problema típico de negociación, donde existe un dominio de la negociación (quién negocia, y sobre qué se negocia), un protocolo de interacción (con qué reglas se resuelve el proceso de negociación), y un conjunto de mecanismos de decisión o *estrategias* que guían a cada uno de los agentes implicados en la negociación a lo largo de cada una de las fases del protocolo de interacción [21].

A. Dominio de la negociación

Para el alcance de este trabajo, vamos a asumir un dominio de negociación multiatributo, donde un acuerdo o solución al problema se define como un conjunto de atributos (*issues*), cada uno de los cuales puede tomar un valor dentro de un determinado rango. En nuestro caso, en un problema de asignación de canales para n_{AP} puntos de acceso, una solución o acuerdo S vendría dada por $S = \{s_i | i \in 1, \dots, n_{AP}\}$, donde cada $s_i \in [1, 11]$ representa la asignación de un canal IEEE 802.11 al punto de acceso i . Aunque en la banda de frecuencia de 2.4GHz se definen 13 canales disponibles, no en todos los lugares del mundo se emplean todos, sino que existen regiones

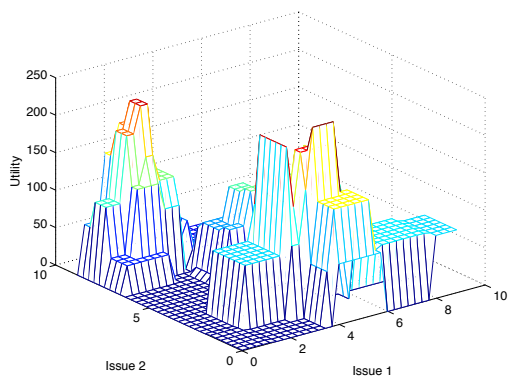


Fig. 2. Función de utilidad con múltiples máximos locales.

como Norteamérica donde se emplean solo los primeros 11 canales. En el presente trabajo, se ha limitado el rango de canales a 11, dado que el cálculo de la interferencia cocanal de la que se hablaba en la sección anterior se ha hecho basándonos en el estudio de [18], en el que se emplean 11 canales únicamente. Pese a que el estándar IEEE 802.11n permite el uso de canales de 40 MHz de ancho de banda (en lugar de los habituales 20 MHz), no se suelen emplear en la banda de 2.4 GHz, debido a la escasez y saturación de los mismos. Por ese motivo, en este trabajo se han considerado únicamente canales de 20 MHz.

Hemos asumido que los puntos de acceso están divididos en dos proveedores (P_1 y P_2), y que cada uno de estos proveedores tiene jurisdicción únicamente sobre la asignación de frecuencias de sus puntos de acceso. Así, P_1 y P_2 serán los dos agentes que negociarán la asignación de frecuencias. Utilizar únicamente dos proveedores en esta primera aproximación nos ha parecido adecuado, puesto que hay mucha más literatura en negociaciones complejas para el caso bilateral que para el multilateral (tres o más agentes).

Finalmente, cada uno de estos dos agentes calculará su utilidad para un determinado acuerdo o solución siguiendo el modelo que se ha presentado en la sección anterior, pero teniendo únicamente en cuenta la interferencia sobre los puntos de acceso que le correspondan. Nuestra hipótesis es que las características del problema (elevado número de soluciones y presencia de interdependencias entre atributos) harán que las funciones de utilidad de los agentes sean altamente complejas, con múltiples máximos y mínimos locales. En la Fig. 2 se muestra un ejemplo de función de utilidad compleja. Desafortunadamente, la dimensionalidad y cardinalidad de los espacios de utilidad que vamos a manejar nos impedirán representarlos gráficamente (o incluso muestrearlos). No obstante, en la sección IV-C veremos resultados que apoyan esta suposición.

B. Protocolo de interacción

Existen multitud de protocolos de interacción para negociaciones bilaterales, desde la clásica alternancia de ofertas [22] a los protocolos basados en subasta [23]. Siguiendo con la hipótesis de que los escenarios de negociación resultantes del problema de asignación de canales en redes IEEE 802.11 serán altamente no lineales, y de acuerdo con la discusión al respecto en [24], hemos empleado un protocolo de mediación

con texto simple (*simple text mediation*) [25]. En su versión más sencilla, el protocolo de negociación funcionaría como sigue:

- 1) Se parte de un primer contrato candidato (S_0^c), generado al azar. Eso supondría, en nuestro caso, asignarle un canal aleatorio a cada punto de acceso.
- 2) En cada iteración t , el mediador propone el contrato S_t^c a los agentes.
- 3) Cada uno de los agentes vota a favor o en contra del contrato propuesto por el mediador.
- 4) El mediador genera un nuevo contrato S_{t+1}^c a partir de los contratos anteriores y de los votos recibidos de los agentes, y se reanuda el proceso en el paso 2.

El proceso continúa hasta que se alcanza un número máximo de iteraciones o se alcanza alguna otra condición de parada. Como puede verse, el protocolo así definido es muy genérico, y debe completarse definiendo los mecanismos de decisión o *estrategias* de los agentes negociadores y del mediador.

C. Mecanismos de decisión

Para el mediador, se ha implementado un mecanismo de generación de nuevos contratos que funciona como sigue:

- Si en un instante t ambos agentes han votado a favor del contrato S_t^c que se les ha presentado, se emplea ese contrato como contrato base S^b para generar el siguiente contrato S_{t+1}^c . En caso contrario, se toma como base el último contrato candidato que haya sido aceptado por ambos agentes.
- Para generar el siguiente contrato candidato S_{t+1}^c se toma el contrato S_b y se cambia el valor de uno de sus atributos al azar. En el caso que nos ocupa, se correspondería con escoger un punto de acceso al azar, y seleccionar de forma aleatoria un nuevo canal para el mismo.
- Después de un número fijo de iteraciones, el mediador anuncia el contrato final, que será el último contrato que haya sido aceptado por los dos agentes.

Para los agentes, se han considerado dos mecanismos diferentes para la emisión de los votos acerca de los contratos candidatos S^c :

- *Hill-climber (HC)*: En este caso, el agente se comporta como un maximizador voraz de utilidad. El agente solo votará a favor de un contrato cuando éste le suponga una ventaja frente al anterior contrato mutuamente aceptado por ambos agentes. En el caso de que no haya contrato previo mutuamente aceptado, el agente votará a favor (esto hace que se acepte siempre el primer contrato propuesto por el mediador).
- *Annealer (SA)*: En este caso se empleará una técnica muy empleada en optimización no lineal conocida como *simulated annealing* [26]. Cuando un contrato suponga una pérdida de utilidad respecto al anterior contrato mutuamente aceptado, existirá una probabilidad de que el agente acepte dicho contrato. Dicha probabilidad P_a depende de la pérdida de utilidad asociada al nuevo contrato Δu y de un parámetro conocido como temperatura de *annealing*, T , de tal forma que $P_a = e^{-\frac{\Delta u}{T}}$. La temperatura parte de un valor inicial, y decrece hasta cero

de forma lineal a lo largo de las sucesivas iteraciones del protocolo.

La elección de estos dos mecanismos no es casual. El uso de *simulated annealing* ha dado muy buenos resultados en negociación para espacios de utilidad no lineales [27], y es la base de varios de nuestros trabajos previos [13], [14]. Por otro lado, tal y como se argumenta en [25], la comparación entre *hill-climbers* y *annealers* permite valorar si el escenario al que nos estamos enfrentando es de alta complejidad, ya que en estos escenarios los negociadores voraces tienden a quedar atrapados en máximos locales, mientras que el negociador basado en *simulated annealing* tiende a escapar de ellos.

IV. EVALUACIÓN EXPERIMENTAL

A. Escenarios considerados

Se han probado diferentes algoritmos de asignación de canales en tres tipos de escenarios. En el primer caso, se han considerado 50 APs y 350 (50·7) clientes, en el segundo 50 APs y 500 (50·10) clientes y en el tercero 100 APs y 500 (100·5) clientes. Tanto la disposición de los APs como de los clientes se ha hecho de manera aleatoria en un plano, de manera que cada cliente se asocia al AP más cercano. Asimismo, si un cliente no tiene ningún AP en su área de cobertura (dada por la sensibilidad de su receptor, tal y como se ha descrito en la Sección II-B), se ha eliminado del problema, algo que también se ha hecho con los APs que no tienen ningún cliente asociado. Obviamente, conforme hay mayor densidad de nodos en la red, son menos los nodos que quedan inconexos. Para cada tipo de escenario y, dada la aleatoriedad de los mismos, se han generado 3 escenarios concretos, por lo que se ha trabajado con un conjunto de 9 escenarios. Además, dada la aleatoriedad de los diferentes algoritmos aplicados, cada uno de los algoritmos se ha ejecutado 10 veces en cada escenario. En la Tabla I se muestra un resumen de los escenarios probados, donde aparecen por un lado los nodos (APs y clientes) inicialmente dispuestos y los nodos que quedan finalmente tras eliminar los inconexos (ν). Además, con el objetivo de cuantificar la densidad de cada escenario, se muestra el número medio de señales interferentes que recibe cada nodo de la red (\bar{I}). Nótese que, aunque parezca un número medio de señales interferentes elevado (se han considerado escenarios con valores de \bar{I} entre 20 y 50), en realidad no lo es tanto, puesto que es habitual, al escanear redes IEEE 802.11, encontrarse conviviendo con 6 u 8 de ellas. Si cada una de ellas tiene 4-6 clientes conectados, alcanzamos un número medio de interferencias entre 24-48. También son habituales escenarios en los que hay pocos APs pero cada uno de ellos alberga a un gran número de usuarios, como por ejemplo en el caso de una red en un campus universitario, donde el AP que dé servicio a un aula tendrá conectado decenas de clientes.

B. Mecanismos analizados

Además de los mecanismos de negociación directamente objeto del estudio, que se han discutido en la sección III-C, se ha realizado una comparación con tres aproximaciones de referencia:

- *Referencia aleatoria*: como una primera línea base de referencia, se ha tomado la elección de una asignación

Tabla I
RESUMEN DE ESCENARIOS EMPLEADOS.

Escenario	# APs	# clientes	ν	\bar{I}
1	50	350	237	22.53
2	50	350	241	21.53
3	50	350	240	21.81
4	50	500	439	34.72
5	50	500	414	39.26
6	50	500	427	34.30
7	100	500	490	47.62
8	100	500	487	51.57
9	100	500	527	48.63

aleatoria de canales. En nuestro contexto, sería el equivalente a que cada punto de acceso escogiese de manera independiente un canal aleatorio en el que transmitir.

- *Sequential Channel Search (SCS)*: una línea base algo más realista, inspirada en LCCS [6]. Se trata de un algoritmo en que los puntos de acceso se activan en secuencia, y cada AP escoge el canal en el que encuentra menos interferencias de otros APs y de clientes pertenecientes a otros APs. Si son varios los canales que se encuentran menos congestionados, se coloca aleatoriamente en uno de ellos.
- *Augmented Lagrangian Harmony Search Optimization (ALHSO)*: Además de nuestros negociadores basados en *simulated annealing*, queríamos tener como referencia un optimizador no lineal independiente que empleara información completa. Hemos escogido *Harmony Search*, que es un algoritmo de optimización evolutivo inspirado en la composición musical [28]. En concreto, hemos usado una implementación disponible públicamente que utiliza multiplicadores aumentados de Lagrange para el manejo de restricciones [29].

C. Resultados

Para la elección de los parámetros de configuración de los escenarios se han escogido valores típicos o razonables desde un punto de vista realista, siendo los siguientes. La potencia de transmisión es de 30 mW, se han considerado antenas básicas con ganancia 0 dB ($G_t = G_r = 0$ dB) y la sensibilidad de los receptores es de -90 dBm. Además, las antenas se encuentran a $h_t = h_r = 1.5$ m de altura sobre el suelo y las pérdidas por obstáculos son de $L = 40$ dB. Todos estos valores sirven para calcular el radio de cobertura, siendo este de $R = 40.3$ m.

En primer lugar se estudian las prestaciones del algoritmo SA con diferentes valores de sus parámetros de configuración: temperatura inicial (T) y número de iteraciones. En general, una mayor temperatura inicial aumenta las posibilidades del optimizador para escapar de máximos locales, aunque una temperatura excesiva puede hacer que el optimizador escape del máximo global, por lo que el ajuste de este parámetro suele ser un proceso delicado y dependiente de la aplicación concreta. Para el caso del número de iteraciones, como regla general un mayor número de iteraciones proporciona más oportunidades de refinar la búsqueda, pero también aumenta el tiempo en que el algoritmo opera con temperaturas elevadas, y por tanto el riesgo de escapar del máximo global (además del coste computacional asociado, por supuesto). En las Fig. 3-5 se muestra cómo varía la utilidad al emplear diferentes valores

de temperatura y de iteraciones para los escenarios 3, 6 y 9 (para el resto de escenarios los resultados son muy similares, por lo que se omiten las gráficas por cuestiones de espacio). En dichas figuras se observa cómo conviene escoger el mayor número de iteraciones posible para obtener mayores valores de utilidad, ocurriendo esto en todos los escenarios estudiados (no únicamente en los mostrados en las figuras), excepto en uno (Fig. 4 con $T = 8$), donde el valor de utilidad desciende de 220.56 a 218.34 al pasar de 2000 a 3000 iteraciones. Respecto a la temperatura, vemos el máximo en $T = 1$. A partir de este momento, el algoritmo SA se ejecutará con $T=1$ y 3000 iteraciones.

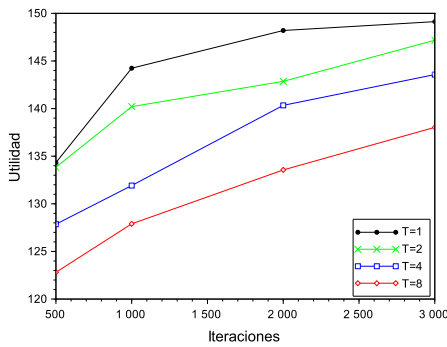


Fig. 3. Evaluación de SA en escenario 3.

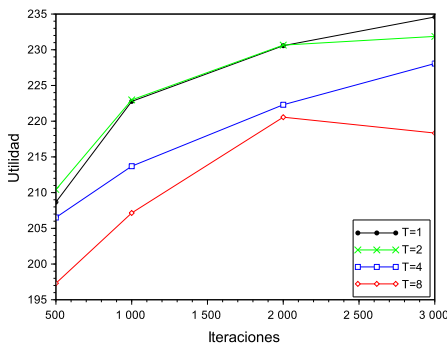


Fig. 4. Evaluación de SA en escenario 6.

A continuación se compara el algoritmo SA con el resto de propuestas: asignación aleatoria de canales, SCS, HC y ALHSO. En las Tablas II-IV se muestra la media y la desviación estándar de la utilidad para las 10 ejecuciones del algoritmo en cada uno de los escenarios. Asimismo, se resalta en negrita la mejor de las soluciones. Como se puede observar, en cuanto a la media de la utilidad, el algoritmo SA se muestra como la mejor solución en todos los casos excepto en uno de ellos (escenario 8), donde ALHSO es la mejor opción. La segunda mejor opción es ALHSO, seguido a cierta distancia de SCS. Como cabía esperar, las peores prestaciones se consiguen con el mecanismo aleatorio, siendo estas muy bajas. En el caso de HC (que puede verse como un optimizador SA con $T = 0$), puede observarse que da buenos resultados en escenarios sencillos (Tabla II), mientras que su

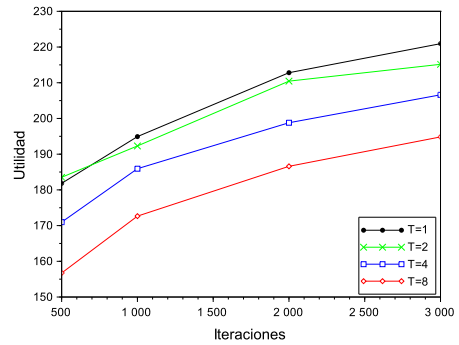


Fig. 5. Evaluación de SA en escenario 9.

Tabla II
UTILIDAD EN ESCENARIOS CON 50 APS Y 350 CLIENTES.

	Escenario 1		Escenario 2		Escenario 3	
	avg	std	avg	std	avg	std
Aleatorio	59.36	9.61	73.74	10.68	65.81	13.76
SCS	106.74	4.30	121.20	7.18	122.78	4.46
ALHSO	121.64	4.17	141.09	3.56	132.92	3.95
HC	123.30	3.50	140.37	4.93	138.02	5.31
SA	130.89	2.65	149.16	1.65	149.13	2.27

rendimiento empeora al movernos a escenarios más complejos (Tablas II-IV), debido a la tendencia a quedarse en máximos locales.

Respecto a la desviación estándar de las soluciones, el algoritmo que presenta una menor desviación estándar es SA en casi todos los casos (excepto en los escenarios más complejos 8 y 9, donde SA es muy similar a ALHSO), seguido de ALHSO, SCS y Aleatorio. Como conclusión, podemos afirmar que superar a un algoritmo de optimización como ALHSO mediante técnicas de negociación como SA, nos indica que este tipo de técnicas son muy válidas y aconsejables para el problema bajo estudio. Puesto que SA requiere, tal y como se ha visto, de la configuración de dos parámetros (temperatura y número de iteraciones), para finalizar se va a hacer una comparativa de todos los algoritmos de manera relativa, con el objetivo de analizar el comportamiento de SA

Tabla III
UTILIDAD EN ESCENARIOS CON 49 APS Y 500 CLIENTES.

	Escenario 4		Escenario 5		Escenario 6	
	avg	std	avg	std	avg	std
Aleatorio	92.96	6.59	78.05	16.95	105.08	13.70
SCS	165.60	9.00	151.74	10.24	181.27	9.14
ALHSO	203.62	5.76	184.01	7.46	221.96	5.37
HC	196.82	9.94	179.11	11.06	211.76	11.05
SA	215.37	2.75	199.52	2.98	234.59	1.46

Tabla IV
UTILIDAD EN ESCENARIOS CON 100 APS Y 500 CLIENTES.

	Escenario 7		Escenario 8		Escenario 9	
	avg	std	avg	std	avg	std
Aleatorio	90.13	9.14	95.73	15.87	88.79	11.64
SCS	164.94	9.34	168.28	13.09	172.70	11.53
ALHSO	216.36	6.40	223.02	5.03	217.53	4.86
HC	196.33	8.07	199.51	8.74	199.65	7.73
SA	220.29	5.17	217.60	5.24	220.94	4.92

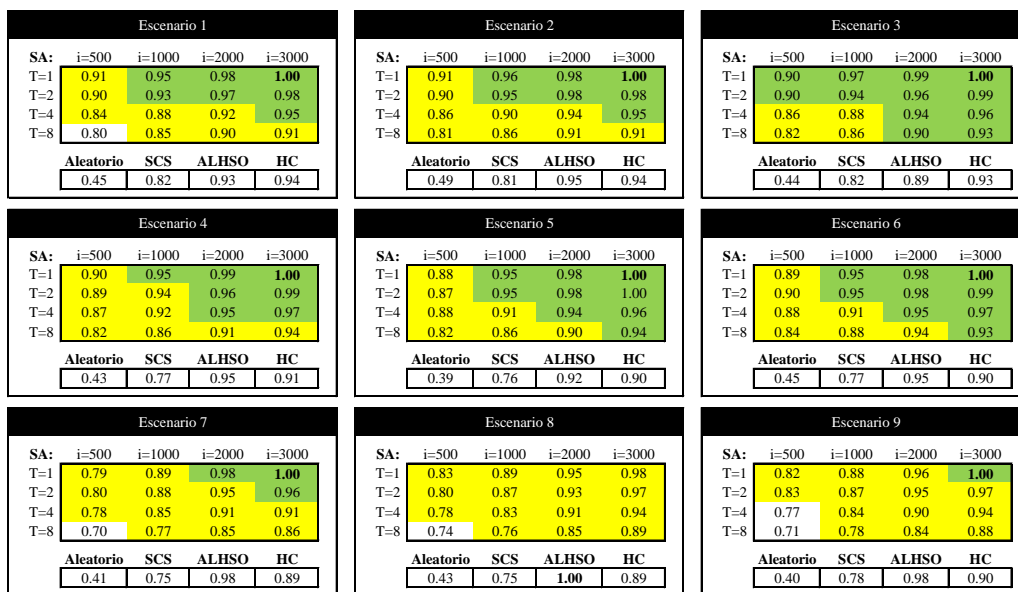


Fig. 6. Resultados relativos al máximo.

si este no se configura adecuadamente.

En la Fig. 6 se muestra el valor medio de la utilidad para los diferentes escenarios y relativo al máximo (en negra). Para facilitar la interpretación, en la tabla de resultados de SA se han marcado en verde las casillas que superan o igualan a ALHSO (también superan a SCS y Aleatorio, pero estos últimos son menos restrictivos). Además, se han marcado en amarillo aquellos casos en los que SA supera a SCS. Así, se observa como para temperaturas cercanas a $T = 1$ y 1000 o más iteraciones, SA supera a ALHSO. Además, en casi todos los casos (excepto para temperaturas altas y muy pocas iteraciones), SA supera a SCS. En la figura se muestran también los resultados de HC para 3000 iteraciones, que equivale a un SA con $T = 0$, y que podemos ver que tiene menor rendimiento que el mecanismo basado en SA para el mismo número de iteraciones.

Como conclusión, se puede recomendar el uso de técnicas de negociación como SA, ya que es mejor en casi todos los casos que SCS y, no es difícil encontrar parámetros de configuración que la hagan mejor incluso que un optimizador no lineal como ALHSO.

V. DISCUSIÓN Y CONCLUSIONES

La asignación de frecuencias en entornos IEEE 802.11 basados en infraestructura no es un tema que haya recibido la suficiente atención por parte de la comunidad investigadora, en relación a su importancia y al impacto que tiene en un número elevado y creciente de usuarios. Esta escasez de trabajos quizás viene dada por la complejidad del problema, que es de tipo NP-complejo [30]. De hecho, en [30] se puede encontrar una revisión de técnicas de asignación de frecuencias en IEEE 802.11, donde se puede observar que no son muchas las propuestas realizadas. A nivel práctico, y hasta donde llega el conocimiento de los autores de esta propuesta, quizás la única técnica real que se emplea es la denominada LCCS (*Least Congested Channel Search*) [6]. Esta técnica, bastante rudimentaria, consiste en que, si un punto de acceso

detecta muchas interferencias en el canal actual, se mueve a la frecuencia menos congestionada en ese momento.

En este artículo se ha modelado el problema de la asignación de canales en redes IEEE 802.11n mediante grafos, de manera que la asignación de canales se traduce en el coloreado de las aristas de un grafo. Puesto que una de las propiedades más características de las redes IEEE 802.11 es que los canales se encuentran parcialmente solapados, se ha incluido dicho efecto en el modelado del problema mediante el uso de una matriz de interferencia cocanal.

Para realizar el coloreado de los grafos se ha planteado el uso de técnicas de negociación automáticas. Más concretamente, se ha empleado un protocolo de negociación bilateral con mediador, en el que los agentes negociadores (dos proveedores de acceso IEEE 802.11, cada uno controlando una fracción de los puntos de acceso del escenario) negocian con una estrategia basada en *Simulated Annealing* (SA). Como referencia para los resultados obtenidos, las aproximaciones basadas en negociación se han comparado con la asignación aleatoria de frecuencias, con una técnica heurística inspirada en LCCS y, finalmente, con la técnica de optimización denominada ALHSO (*Augmented Lagrangian Harmony Search*), que se basa en un algoritmo evolutivo inspirado en la música y que es muy adecuado para problemas de optimización con espacios de soluciones no lineales. Los resultados muestran que las prestaciones de la técnica aleatoria son muy pobres, seguida por SCS, que presenta prestaciones bastante superiores. Los mejores resultados (tanto para valores medios de la utilidad como para desviaciones estándar de la misma) se obtienen con las técnicas de negociación utilizando SA. Además, se han probado diversas configuraciones para el algoritmo SA, encontrándose prestaciones superiores de SA respecto a ALHSO en un buen número de configuraciones. Como conclusión general, se puede decir que las técnicas de negociación se muestran como muy útiles para la resolución de problemas de asignación de canales a redes IEEE 802.11.

Respecto a trabajos relacionados con nuestra propuesta, cabe citar los trabajos de [31] y [32], aunque no se centran en IEEE 802.11 ni emplean técnicas de negociación. Una propuesta que trata de coordinar puntos de acceso IEEE 802.11 para la asignación de frecuencias sin el empleo de grafos se puede encontrar en [33]. Aunque quizá el trabajo más relacionado con nuestra propuesta sea el de [34], por estar inspirado en el problema del coloreado de grafos y por utilizar una medida de la bondad de las soluciones obtenidas durante el proceso de optimización que recuerda mucho a las funciones de utilidad que se emplean en negociación.

Aunque los resultados de los experimentos son satisfactorios, quedan muchas posibilidades de trabajo futuro. Una vez demostrado que las técnicas de negociación no lineal pueden aplicarse satisfactoriamente a este sistema, cabe plantearse una comparativa entre una selección más amplia de las mismas. También estamos interesados en estudiar la influencia de diferentes propiedades de los escenarios en la adecuación de unas y otras (de un modo similar a como planteamos en [35]). Finalmente, sería interesante analizar como escalan este tipo de técnicas no sólo con el tamaño de los escenarios, sino también con el número de proveedores (incluso llegando al caso límite de APs independientes).

Como trabajo futuro se plantea, en primer lugar, implementar otras técnicas de negociación automática (incluidas multilaterales) que pueden ser útiles para resolver el problema, además de otras técnicas heurísticas que se han propuesto en la literatura. Por otro lado, se quiere probar dichas técnicas sobre grafos que reflejen un escenario real tridimensional.

AGRADECIMIENTOS

Este trabajo ha sido financiado por los proyectos TIGRE5-CM Comunidad de Madrid (S2013/ICE-2919), ATLAS-Project VA172A12-2, MICINN MTM2011-22792, MINECO MTM2014-54207 y MINECO TIN2014-61627-EXP.

REFERENCIAS

- [1] E. Au, M. Cheong, C. Ngo, C. Cordeiro, W. Zhuang, "The future of Wi-Fi - Guest Editorial", IEEE Communications Magazine, vol.52, no.11, pp.20,21, 2014.
- [2] P. Dalal, M. Sarkar, K. Dasgupta, N. Kothar, "Link Layer Correction Techniques and Impact on TCP's Performance in IEEE 802.11 Wireless Network", Communications and Network, vol.6, no.2, ID 45714, 2014.
- [3] R. Liao, B. Bellalta, M. Oliver, Z. Niu, "MU-MIMO MAC Protocols for Wireless Local Area Networks: A Survey," IEEE Communications Surveys & Tutorials (pendiente de publicación).
- [4] K. Aardal, S. van Hoesel, A. Koster, C. Mannino, A. Sassano, "Models and solution techniques for frequency assignment problems", Annals of Operations Research, vol. 153, no. 1, pp. 79-129, 2007.
- [5] F. Luna, E. Alba, Antonio J. Nebro, S. Pedraza. "Evolutionary algorithms for real-world instances of the automatic frequency planning problem in GSM networks." Evolutionary Computation in Combinatorial Optimization, pp. 108-120. Springer Berlin Heidelberg, 2007.
- [6] M. Achanta, "Method and apparatus for least congested channel scan for wireless access points", US Patent App. 10/959,446, 2006.
- [7] F. Ren, M. Zhang, K. M. Sim, "Adaptive conceding strategies for automated trading agents in dynamic, open markets" Decision Support Systems, vol.46, no.3, pp.704-716, 2009.
- [8] R. B. Myerson, M. A. Satterthwaite, "Efficient mechanisms for bilateral trading", Journal of Economic Theory, vol. 29, no. 2, pp. 265-281, 1983.
- [9] T. Ito, M. Klein, H. Hattori, "A multi-issue negotiation protocol among agents with nonlinear utility functions", Multiagent Grid Syst., vol.4, no.1, pp.67-83, 2008.
- [10] K. M. Sim, B. Shi, "Concurrent negotiation and coordination for grid resource coallocation", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics - Special issue on game theory, vol.40, no.3, pp.753-766, 2010.
- [11] R. G. Smith, "The contract net protocol: High-level communication and control in a distributed problem solver", IEEE Transactions on Computers, vol. C-29, no.12, pp.1104-1113, 1980.
- [12] M. Wu, M. de Weerd, H. La Poutré, "Efficient methods for multi-agent multi-issue negotiation: Allocating resources", Principles of Practice in Multi-Agent Systems, Lecture Notes in Computer Science vol. 5925, pp. 97-112, 2009.
- [13] I. Marsa-Maestre, M. A. Lopez-Carmona, J. R. Velasco, E. de la Hoz, "Avoiding the prisoner's dilemma in auction-based negotiations for highly rugged utility spaces", En Actas de 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), vol.1, no.1, pp.425-432, 2010.
- [14] I. Marsa-Maestre, M. A. Lopez-Carmona, J. R. Velasco, T. Ito, M. Klein, K. Fujita, "Balancing utility and deal probability for auction-based negotiations in highly nonlinear utility spaces", En Actas de 21st International Joint Conference on Artificial Intelligence (IJCAI), pp. 214-219, 2009.
- [15] Z. Tuza, Graph coloring, *Handbook of Graph Theory*, Discrete Mathematics and Its Applications, vol. 25, pp. 408-438, CRC Press, 2003.
- [16] E. Z. Tragos, S. Zeadally, A. G. Fragkiadakis, V. A. Siris, "Spectrum Assignment in Cognitive Radio Networks: A Comprehensive Survey", IEEE Comm. Surveys Tutorials, vol. 15, no 3, pp. 1108-1135, 2013.
- [17] D.B. Green, A.S. Obaidat, "An accurate line of sight propagation performance model for ad-hoc 802.11 wireless LAN (WLAN) devices", En Actas de IEEE International Conference on Communications (ICC), vol. 5, pp. 3424-3428, 2002.
- [18] S.W.K. Ng, T.H. Szymanski, "Interference measurements in an 802.11n wireless mesh network testbed", En Actas de 25th IEEE Canadian Conference on Electrical and Computer Engineering, 2012.
- [19] A. Bazzi, "On uncoordinated multi user multi RAT combining", In Proceedings of IEEE Vehicular Technology Conference Fall, 2011.
- [20] J. Geier, "How to: Define Minimum SNR Values for Signal Coverage" Wireless-Nets Independent Consulting. http://www.wireless-nets.com/resources/tutorials/define_SNR_values.html (acc. 1/4/2015).
- [21] S. Fatima, M. Wooldridge, N. R. Jennings, "Optimal negotiation strategies for agents with incomplete information, Lecture Notes on Computer Science, vol. 2333, pp. 53-68, 2001.
- [22] A. Rubinstein, "Perfect equilibrium in a bargaining model", *Econometrica*, vol. 50, no. 1, pp. 97-109, 1982.
- [23] H. Hattori, M. Klein, T. Ito, "Using iterative narrowing to enable multi-party negotiations with multiple interdependent issues", En Actas de 6th international joint conference on Autonomous agents and multiagent systems (AAMAS), pp 1-3, 2007.
- [24] I. Marsa-Maestre, "Contribución a la negociación automática en espacios de utilidad complejos", PhD thesis, Universidad de Alcalá, 2009.
- [25] M. Klein, P. Faratin, H. Sayama, Y. Bar-Yam, "Negotiating complex contracts", Group Decision and Negotiation, vol. 12, no. 2, pp. 111-125, 2003.
- [26] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by simulated annealing", *Science*, vol.220, no.4598, pp.671-680, 1983.
- [27] F. Lang, A. Fink, "Learning from the metaheuristics: Protocols for automated negotiations", Group Decision and Negotiation, vol. 24, no.2, pp.299-332, 2015.
- [28] Z.W. Geem, J.H. Kim, G.V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search", *Simulation*, vol. 76, no.2, pp. 60-68, 2001.
- [29] R. E. Perez, P. W. Jansen, J.R.R.A. Martins, "pyOpt: A Python-Based Object-Oriented Framework for Nonlinear Constrained Optimization", Structures and Multidisciplinary Optimization, vol.45, no.1, pp. 101-118, 2012. <http://www.pyopt.org/reference/optimizers.alhso.html>.
- [30] S. Chiochan, E. Hossain, J. Diamond, "Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey", IEEE Communications Surveys Tutorials, 2010.
- [31] C. McDiarmid, "Frequency-distance constraints with large distances", *Discrete Mathematics*, vol. 223, no. 1-3, pp. 227-251, 2000.
- [32] L. Narayanan, S.M. Shende, "Static frequency assignment in cellular networks", *Algorithmica*, vol. 29, no. 3, pp. 396-409, 2001.
- [33] M. Abusubaih, J. Gross, S. Wiethoelter, A. Wolisz, "On access point selection in IEEE 802.11 wireless local area networks", En Actas de 31st IEEE Conference on Local Computer Networks, pp 879-886, 2006.
- [34] A. Mishra, S. Banerjee, W. Arbaugh, "Weighted Coloring Based Channel Assignment for WLANs", ACM SIGMOBILE Mobile Computing and Communications Review, vol. 9, no 3, pp. 19-31, 2005.
- [35] I. Marsa-Maestre, M. Klein, C. M. Jonker, R. Aydoğan, "From problems to protocols: Towards a negotiation handbook", Decision Support Systems, vol. 60, no.0, pp.39-54, 2014.

Estimación Eficiente del RSSI en Redes WiFi para Servicios Inalámbricos Futuros

José A. Santana^{1,3}, Elsa Macías López^{2,3}, Vicente Mena^{1,3}, Domingo Marrero^{2,3}, Álvaro Suárez^{2,3}

¹Departamento de Señales y Comunicaciones
{jose.santana,vicenteeingenio.mena}@ulpgc.es

²Departamento de Ingeniería Telemática
{alvaro.suarez, elsa.macias, domingo.marrero}@ulpgc.es

³Grupo de Arquitectura y Concurrencia (GAC)
Universidad de Las Palmas de Gran Canaria
Campus Universitario de Tafira s/n
35017 Las Palmas de Gran Canaria

Resumen- El RSSI es una escala de referencia que indica la fuerza de la señal de radio (en relación a 1 mW) recibida en un terminal inalámbrico o móvil. Desde hace años este parámetro ha sido utilizado para iniciar distintos servicios de redes móviles e inalámbricas como por el ejemplo la fase de *scanning* en el *handoff*. A lo largo de los años ha sido empleado y desestimado por no ser útil para esos procedimientos debido a su elevada variabilidad a lo largo del tiempo (*volatilidad*), imposible de ser estimada con precisión. Pero, recientemente, un artículo publicado en una revista de elevado impacto ha probado matemáticamente, para señales de laboratorio y reales, que es posible usarlo. Y demuestra que, en el dominio del tiempo, este parámetro tiende siempre a retornar a medias móviles. Esto implica que se puede usar una estimación precisa del RSSI para mejorar las aplicaciones de las redes inalámbricas actuales, e incluso mejorar las nuevas generaciones de tecnologías de radio (5G). En este artículo nosotros proponemos un método novedoso que estima las medias móviles del RSSI, mejorando considerablemente un método propio anterior y los métodos que se han propuesto en este año y el pasado. Los resultados experimentales así lo demuestran. Además, presentamos distintas aplicaciones telemáticas que mejoran su rendimiento aplicando nuestro método.

Palabras Clave- RSSI, Wireless Fidelity, Estimación, Medias Móviles, Aplicaciones Inalámbricas.

I. INTRODUCCIÓN

Es ampliamente conocido que uno de los problemas típicos que experimentan los usuarios de *Wireless Fidelity* (WiFi) es la desconexión repentina e intermitente de su terminal inalámbrico, sin que en principio haya razón para ello, porque el terminal muestra un buen nivel de cobertura radio. Nosotros hemos estudiado ampliamente este fenómeno y presentado algunos métodos para mitigar su efecto en servicios multimedia [1]. Con el *Received Signal Strength Indicator* (RSSI) se puede estimar el instante de tiempo aproximado en el que podría ocurrir una desconexión, a partir de las medidas de referencia que se obtienen de las balizas (emitidas en el instante indicado en su campo *TimeStamp* (TS)) que los *Access Points* (AP) emiten regularmente cada *Beacon Interval* (BI). Pero tiene otras

muchas aplicaciones (incluso comerciales) importantes: como por ejemplo la que pretende explotar *Google* en [2].

Históricamente el RSSI ha sido criticado porque no proporciona una buena estimación del nivel de fuerza de la señal radio, debido a que presenta una variabilidad, a lo largo del tiempo, elevada e impredecible. Los motivos [3], entre otros son: disparidad en la escala de medida según el hardware/software, interferencias radioeléctricas en la banda espectral utilizada, y retraso de transmisión de las balizas debido a incidencias del algoritmo de control de acceso al canal. Sin embargo, nosotros mostramos, en un trabajo pionero [4], que en la práctica se puede hacer estimaciones precisas de RSSI, debido a que su tendencia natural es a retornar siempre a una media móvil. Pero no lo demostramos matemáticamente con la suficiente rigurosidad. Este trabajo ha sido mejorado recientemente y publicado en una revista de elevado impacto [5] demostrando, además, que un método basado en ventanas puede mejorar la estimación. Pero tampoco demuestra formalmente el retorno a las medias móviles del RSSI a lo largo del tiempo. Esa demostración formal se ha realizado recientemente en una publicación de elevado impacto [6]. En la misma se determina que el comportamiento típico del RSSI a lo largo del tiempo se caracteriza por una secuencia de *saltos* (*jumps*), causados por efectos inherentes del canal, que regularmente desvían su tendencia natural de retorno a las medias móviles. Cuando hay saltos esporádicos puntuales no afecta a la estimación del RSSI; pero, si se concentran en un intervalo de tiempo, pueden modificar la media móvil. Por tanto, estimar las medias móviles del RSSI, teniendo en cuenta estos factores, es un tema de suma importancia, debido a las aplicaciones que se derivan de una buena estimación en las redes inalámbricas (localización, *handover*, asociación eficiente a AP...). En ese trabajo se deja claro que el RSSI es adecuado para iniciar los procesos de *handover* en redes inalámbricas. Sin embargo, ni en ese trabajo ni anteriores se tiene en cuenta que la ocupación desproporcionada del canal inalámbrico, o bien la ejecución de servicios de aplicación en el terminal inalámbrico (entre otras) pueden provocar una alteración de los *Target Beacon Transmission Time* (TBTT) que el AP debe transmitir separados por un BI [7], generándose *Beacon*

Delay (BD), respecto al instante de transmisión esperado, con jitter. Esto tiene una importancia significativa porque, en caso de que el BD sea elevado, los métodos anteriores asignan valores promediados de RSSI en intervalos de tiempo en los que originalmente no hay medida de RSSI (sample). Con esto se produce una tergiversación que influye al nivel de aplicación, porque ésta puede recibir una estimación del RSSI y actuar erróneamente en consecuencia (por ejemplo decidir hacer un proceso de handover), cuando en realidad la medida del RSSI no existía.

En este artículo presentamos un método novedoso de estimación eficiente de la tendencia del RSSI a las medias móviles a lo largo del tiempo, definiendo ventanas de medidas de RSSI de tamaño variable, y aplicando una modificación de nuestro método anterior dentro de la ventana de medidas de RSSI que tienden a una media móvil. Nuestro método es el primero en definir, automáticamente, ventanas de clasificación del RSSI que evitan la estimación del siguiente valor cuando la nueva medida captada provoca un BD elevado, o una variación brusca de amplitud del RSSI. Con ello se minimiza el efecto de estas medidas en el error de estimación, se optimiza la adaptación de ésta a las medias móviles, y se proporciona información más detallada, que otros métodos, al nivel de aplicación. Los resultados experimentales demuestran que nuestra estimación mejora considerablemente la de otros métodos.

En el apartado II presentamos nuestro procedimiento comparándolo con los otros. En el apartado III presentamos resultados experimentales comparados con los de otras técnicas. En el apartado IV analizamos las aplicaciones de nuestro algoritmo en redes inalámbricas y, finalmente presentamos algunas conclusiones.

II. EL MÉTODO PROPUESTO

El principal objetivo del método que proponemos es segmentar la evolución del RSSI en ventanas de tamaño variable a lo largo del tiempo, para tratar de capturar la tendencia a las medias móviles de esa evolución temporal y, por otro lado, no estimar el RSSI cuando se produce un BD elevado, o una variación brusca de amplitud del RSSI. Para ello, en primer lugar, definimos el concepto *Uncertainty Interval (UI)* como el periodo de tiempo entre la llegada de dos balizas consecutivas, en el cual no tenemos certeza absoluta de cómo ha variado el RSSI. Pronosticar que la estimación sigue la tendencia a la media móvil entre ambas medidas es sólo una de las opciones que se puede asumir. Clasificamos el UI en dos tipos:

- *Short Uncertainty Interval (SUI)*. Un UI de corta duración que denota la cadencia esperada de llegada de balizas. En general se cumple que $0 \leq SUI \leq BI$.
- *Long Uncertainty Interval (LUI)*. Un UI prolongado que se produce cuando $BD \gg BI$, provocado por incidencias en el canal u otra causa, y en general impredecible en tiempo real. Estimar el RSSI en un LUI no sería adecuado. Proponemos aislar dicho intervalo para que sea el nivel de aplicación el que determine qué acciones realizar, ya que éste puede conocer más información acerca del funcionamiento de los protocolos de nivel de enlace, red y de transporte usando técnicas *Cross-layer* [8].

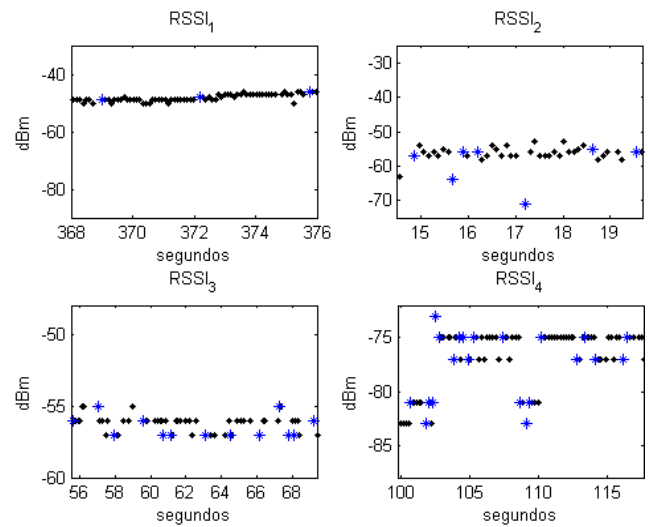


Fig. 1. Efecto del LUI en la evolución temporal de RSSI tomadas para hacer pruebas experimentales.

Para ilustrar estos conceptos, en la Fig. 1 se muestra un tramo de cuatro de las señales RSSI (denominadas $RSSI_1$, $RSSI_2$, $RSSI_3$ y $RSSI_4$) que se utilizaron para probar nuestro método. Los puntos negros son los valores reales de RSSI capturados frente a un eje de tiempos en segundos. Cuando se produce un cambio significativo de amplitud de RSSI o la siguiente baliza se demora demasiado en llegar, se produce un cambio brusco en la dispersión de las muestras. Todos los métodos de estimación del RSSI analizados suavizan o eliminan dichas variaciones bruscas. Nosotros proponemos que, esas medidas que se salen del rango medio de amplitud de la ventana actual, sean tratadas en otras ventanas. Esto es, nuestro método señala esta situación con un cambio de ventana (instantes marcados con un asterisco azul en las gráficas). Y, en el caso de que también se salgan del rango del resto de las medidas de la ventana siguiente, se tratan como medida única de la ventana intermedia. Esto equivaldría al concepto de *Agujero* de [4,5] pero extendido a cambios bruscos positivos de potencia, y sin asignarle un valor concreto. Quedaría al arbitrio de la aplicación concreta el tratamiento de estas ventanas de una sola muestra.

A. Formalización de la estimación de las medias móviles

Sea $V=[I, K]$ la progresión aritmética de paso unitario que representa el dominio de definición del RSSI (X_k) asociado a la baliza que llegó en el instante TS_k :

$$X_k: \mathbb{N}^* \rightarrow \mathbb{Z} \quad (1)$$

$$k \rightarrow X_k, \quad k \in V = [1, K] \subset \mathbb{N}^*$$

Además, aunque el AP intenta transmitir balizas cada BI, que suele ser de 0.1024 s, el BD real de la k -ésima baliza, que nosotros asociamos a X_k , se obtiene como:

$$BD_k = TS_k - TS_{k-1} \quad (2)$$

Así, aunque el valor real de RSSI entre TS_k y TS_{k+1} es desconocido, podemos asumir que:

- Para un SUI es razonable que la estimación se ajuste a la media móvil. Por tanto el error de aproximación de la estimación sería tolerable a nivel de aplicación.

- Para un LUI no se puede asumir que exista una tendencia a la media móvil, porque este intervalo puede ser muy prolongado, e incluso podríamos encontrarnos ante un periodo de ausencia de detección de señal radio en el terminal. Cualquier estimación en un LUI, como hacen los demás autores, sería desacertada. No se debe estimar en este caso el RSSI.

Conocida la medida (TS_k, X_k) , que representa el instante TS_k en que llegó la k -ésima baliza y su RSSI asociado, definimos la estimación (TS_k+BI, XP_k) como el pronóstico del instante de llegada de la siguiente baliza y su RSSI. Estas estimaciones las calculamos a partir de ventanas de medidas de RSSI $\{X_{k-N_m+1}, X_{k-N_m+2}, \dots, X_k\}$, donde N_m , el tamaño de la m -ésima ventana V_m (partición de V), es variable. De tal forma que son dos las condiciones necesarias para incluir (TS_k, X_k) en la m -ésima ventana V_m :

1. Similitud temporal ($BD_k \leq SUI$).
2. Similitud de RSSI, entendiendo por ésta que los X_k de la ventana no presenten variaciones bruscas respecto de un umbral establecido.

Mediante un proceso iterativo se construye un número variable M de ventanas disjuntas V_m , que forman parte del conjunto V total de medidas X_k , siendo K el cardinal de V :

$$V = \bigcup_{m=1}^M V_m \wedge \bigcap_{m=1}^M V_m = \emptyset, \quad V = [1, K] \subset \mathbb{N}^* \quad (3)$$

Definimos *Acceptable Uncertainty Window* (V_m) como una partición disjunta de N_m muestras consecutivas de V , que podría expresarse como un conjunto de progresiones aritméticas de paso unitario no solapadas:

$$V_1 = [1, N_1], \quad V_m = [N_{m-1} + 1, N_{m-1} + N_m] \quad (4)$$

$$K = \sum_{m=1}^M N_m \quad (5)$$

Denominamos *Window Transition* (WT) al primer elemento de cada ventana V_m , separado de la ventana anterior V_{m-1} por una disimilitud considerable en tiempo o en amplitud de RSSI. Constituyen un concepto equivalente al de *Agujero* en [4] o *salto* en [6]. La medida (TS_k, X_k) que lo originó debe ser ignorada, aunque su tratamiento concreto se deja a la aplicación que utilice este método de estimación:

$$\forall m \in [1, M]: WT_m = V_m(1) \quad (6)$$

El criterio para realizar una WT debe preservar nuestro objetivo de mantener una coherencia temporal y de amplitud entre todas las medidas de RSSI de la ventana, a la vez que acercarnos a una precisión en la estimación preestablecida.

$$k \in V_m \Leftrightarrow BD_k < u \wedge |X_k - XP_k| < \varepsilon \quad (7)$$

Donde el umbral u debería estar en función del promedio de BD del AP, ya que según [7] el AP puede demorarse en su envío por incidencias del canal o por encontrarse ocupado en tareas de alto nivel. Además, en [9] se demuestra que el retraso de balizas se ajusta a una distribución Gamma o Logística. Las curvas de estas distribuciones son muy similares a la mitad derecha de una distribución normal. Parece adecuado entonces utilizar la regla empírica en base al BD medio para determinar el BD máximo aceptable, antes de cambiar de ventana. La media de la distribución se recalcula a la llegada de cada baliza.

$$k = 1, \quad m = 1, \quad N_m = 1, \quad V_m = \{k\}, \quad \hat{a}_m = 0$$

$$BD_k = 0, \quad error = 0, \quad u = BI, \quad S = BI^2$$

$$ty = TS_k + BI, \quad Y = X_k$$

Para cada baliza k

Si $|BD_k - u| \leq 3 \cdot \sqrt{S} \wedge error \leq \varepsilon$ Entonces

$$t_k = ty, \quad XP_k = Y$$

$$N_m = N_m + 1, \quad V_m = \{V_m, k\}$$

Sino

$$m = m + 1, \quad N_m = 1, \quad V_m = \{k\}, \quad \hat{a}_m = 0$$

$$t_k = TS_k + BI, \quad XP_k = X_k$$

Fin Si

$$k = k + 1$$

$$BD_k = TS_k - TS_{k-1}$$

$$u = (u \cdot (k - 1) + BD_k) / k$$

$$S = (S \cdot (k - 1) + BD_k^2) / k$$

$$a_k = (X_k - X_{k-1}) / (TS_k - TS_{k-1})$$

$$\hat{a}_m = (\hat{a}_m \cdot N_m + a_k) / (N_m + 1)$$

$$ty = TS_k + BI, \quad Y = \hat{a}_m \cdot (ty - TS_{WT_m}) + X_{WT_m}$$

$$error = |X_k - Y|$$

Fin Para

Fig. 2. Algoritmo de estimación de RSSI con ventanas adaptativas.

Por otro lado, el umbral ε debe tener en cuenta el nivel de cuantificación de la RSSI. Como X_k es entero, una precisión de la predicción XP_k del orden de $\varepsilon=0.5$ resulta adecuada, siendo igualmente correcta cualquier precisión $0 < \varepsilon < 1$.

Assumiendo que, dentro de la ventana V_m , la evolución de la señal de potencia seguiría una tendencia coherente, el método del Gradiente ajustaría mejor a la misma su estimación de la X_k , $k \in V_m$ del XP_k en el instante TS_k+BI :

$$a_k = \frac{X_k - X_{k-1}}{TS_k - TS_{k-1}}, \quad \forall k \in [1, K] \quad (8)$$

$$\hat{a}_m = \frac{\sum_{k \in V_m} (a_k)}{N_m} = \frac{\sum_{k \in V_m} \left(\frac{X_k - X_{k-1}}{TS_k - TS_{k-1}} \right)}{N_m}, \quad \forall m \in [1, M] \quad (9)$$

$$XP_{k+1} = \hat{a}_m \cdot (TS_k + BI - TS_{WT_m}) + X_{WT_m} \quad (10)$$

A partir de este modelado del cálculo del XP_k , se construye un procedimiento de baja complejidad algorítmica. En él básicamente se implanta una simplificación de Ec. 9, que permite sustituir la suma de N_m fracciones en cada paso por una suma ponderada de solo dos términos: la fracción actual (Ec. 8) y el término \hat{a}_m del paso anterior.

Los pasos que debe dar el terminal inalámbrico para estimar, según nuestro método, el siguiente valor de RSSI se muestran en la Fig. 2. Básicamente consiste en un bucle principal que espera la llegada de la siguiente baliza. Una vez recibida se decide si su X_k debe incluirse en la ventana actual, dependiendo de que se haya recibido tras un LUI y/o se haya producido un desvío importante del valor de amplitud de

RSSI esperado. Como paso final, mediante una sencilla operación, se actualiza la pendiente promedio de la ventana, utilizando la contribución del último X_k recibido. El error de aproximación y el retraso de baliza influyen en el criterio de construcción de las ventanas en la siguiente iteración.

B. Comparativa con métodos similares

Numerosos autores [3-6,10,11] han desarrollado algoritmos para anticipar la evolución del RSSI emitido por los APs. Analizamos algunos de los más eficientes, haciendo especial hincapié en los que consiguen aproximar su estimación a las medias móviles con el mínimo error.

Aunque nosotros en [4] no definíamos formalmente el retorno del RSSI (que en [4] expresamos como x_k) a las medias móviles a lo largo del tiempo, sí que la asumíamos implícitamente en nuestro método de estimación. Se trata de un procedimiento de muy baja complejidad algorítmica en el que se promedian las pendientes de todas las parejas x_k y x_{k+1} . En esta aproximación se asume que el BD_k puede variar; pero no se explota la localidad de la tendencia del RSSI a una media móvil particular, tratando todas las x_k en una única ventana. En cuanto a las variaciones bruscas de amplitud del RSSI, sólo teníamos en cuenta su descenso de valor, y los denominamos *agujero* de cobertura. Su tratamiento consistía en filtrarlos asignándoles el valor mínimo, que es cero tras re-escalar previamente todos los x_k al intervalo 0%-100%.

Así, conocida las medidas de RSSI hasta x_{k-1} y sus instantes de captura, se estima el siguiente RSSI (\hat{x}_k) para un instante prefijado t_k , de acuerdo con las Ec. 11-13:

$$\hat{x}_k^- = x_0 + \frac{t_k - t_0}{k-1} \cdot \sum_{n=1}^{k-1} \left(\frac{x_k - x_{k-1}}{t_k - t_{k-1}} \right) \quad (11)$$

$$\sigma(x_{k-1}, \hat{x}_k^-) = \sqrt{\frac{1}{2} \cdot (x_{k-1} - \hat{x}_k^-)^2} \quad (12)$$

$$\hat{x}_k = \hat{x}_k^- + S(x_{k-1} - \hat{x}_k^-) \cdot \sigma(x_{k-1}, \hat{x}_k^-) \cdot S(x_{k-1} \cdot x_k)^2 \quad (13)$$

Concretamente en la Ec. 11 se calcula el promedio de pendientes hasta el instante actual. Debido al uso de todas las medidas de RSSI (hasta x_{k-1}), la aproximación pierde precisión en su adaptación en zonas localizadas. La Ec. 12 es un término corrector que trata de ajustar la estimación al valor de la x_{k-1} . Esta compensación se integra en la fórmula general descrita en la Ec. 13. En ésta S representa la función signo, con lo cual la compensación se suma o resta de acuerdo al signo de la propia compensación. Y el último factor $S(x_{k-1} \cdot x_k)^2$ simplemente anula o activa la compensación dependiendo de que la muestra actual o la anterior tengan valor 0% (agujero). Este esquema es algo limitado porque no considera los casos en que un x_k es pequeño no nulo, o significativamente alto respecto al x_{k-1} .

En este mismo trabajo estudiamos el rendimiento del filtro de Kalman, pero también llegamos a la conclusión de que su eficiencia es muy baja.

El trabajo de *Gradiente Mejorado* [5] también se basa en nuestro método de *Gradiente*. Su principal aportación es que adaptan el algoritmo para promediar la pendiente de solo una ventana prefijada de k muestras en cada iteración, como se expresa en la Ec. 14:

$$X_n = \left(\frac{1}{k} \cdot \sum_{i=n-k}^{n-1} \left(\frac{x_i - x_{i-1}}{t_i - t_{i-1}} \cdot (t_n - t_{n-1}) \right) \right) + x_{n-1} \quad (14)$$

Este tamaño fijo en la ventana, en lugar de un tamaño adaptativo como sugerimos en el presente artículo, es una de las causas de que *Gradiente Mejorado* genere mayores errores de estimación que nuestro método. Además de utilizar una ventana de ancho fijo, incurre en un grave error al utilizar en su método el instante real asociado a \hat{x}_k . Debemos ser conscientes que, al igual que desconocemos el valor real de X_{k+1} , también desconocemos TS_{k+1} . Éste incluso podría demorarse significativamente, haciendo cambiar considerablemente el ajuste de la curva estimada. Además, su algoritmo de cálculo del gradiente medio implica sumar N fracciones en cada paso, cuando la simplificación que nosotros proponemos en nuestro actual trabajo la convierte en una única suma ponderada de una fracción con el término anterior.

En la Ec. 15, definen un factor de compensación análogo al de la Ec. 12.

$$\sigma(X_n, x_n) = \sqrt{\frac{1}{2} \cdot (X_n - x_n)^2} \quad (15)$$

Dicho factor se utiliza en su fórmula general de predicción de x_f mostrada en la Ec. 16 para corregir la aproximación a la muestra anterior x_n , de forma similar a como se hacía en la Ec. 13. Pero, a diferencia del método de *Gradiente*, en la Ec. 16 la detección de agujeros se implementa volviendo a multiplicar por el cuadrado del mismo factor de corrección:

$$x_f = X_n + S(X_n - x_n) \cdot \sigma(X_n, x_n) \cdot \sigma(X_n, x_n)^2 \quad (16)$$

No queda claro que este detector de agujeros vaya a tener la misma eficiencia que el cuadrado del signo del producto de la muestra actual por la anterior $S(X_n \cdot x_n)^2$, como usamos nosotros en el método *Gradiente*. Por otro lado, en el *Gradiente Mejorado* las muestras RSSI no se re-escalan al rango 0%-100%, por lo que los agujeros aparecen como máximos en la señal estimada, lo que no consideramos adecuado. Finalmente, este método sigue adoleciendo de un criterio que filtre también los cambios bruscos positivos de medidas de RSSI.

El trabajo presentado en [6] fue el primero que demostró formalmente esa tendencia del RSSI a las medidas móviles a lo largo del tiempo. Además, aprovecha la variación local de los X_k agrupándolos en ventanas, aunque de tamaño fijo. Una de las desventajas de su método de estimación, además del costoso ajuste de regresión, es que además precisa de un suavizado previo de los X_k . Y, a nivel de precisión, supone que todos los X_k se reciben siempre a un BD constante que, como se ha demostrado, no es cierto siempre (y muy poco probable en general). Si el BD es elevado, la curva estimada no se ajusta adecuadamente a la separación entre muestras. La forma en que tratan los X_k que se desvían de la tendencia a la media móvil (que llaman *saltos*) requiere un elevado número de iteraciones en cada ventana, y por tanto implica un elevado coste computacional. Además, integran parámetros aleatorios como una parte más del modelo de estimación, lo que complica su algoritmo (sin quedar claro que contribuya a su funcionamiento general).

Este último método fue mejorado por nuestro trabajo en [3], en el que generamos una estimación más aproximada y

estable, eliminando los términos aleatorios citados, aunque manteniendo el costoso bucle de tratamiento de saltos.

El modelo que ahora proponemos construye ventanas de medidas de RSSI correladas en amplitud/tiempo, por lo que se tiende de forma natural a una media móvil dentro de una ventana. Además, pretende minimizar las distancias entre la trayectoria formada por la secuencia de medidas (TS_k, X_k) a lo largo del tiempo y la constituida por la sucesión de sus estimaciones (TS_k+BI, XP_k). Para ello agrupamos en ventanas solo las medidas (TS_k, X_k) que presenten una estrecha similitud entre la anterior y la siguiente. Este planteamiento es equivalente a la estrategia del modelo de [3] de eliminar los saltos que dificultan obtener un *scatter plot* de pares ordenados (X_k, X_{k+1}) con una pendiente de 45°. Las diferencias son que en este modelo buscamos la similitud ventana a ventana, que la temporización afecta a la segmentación mediante ventanas, y que no precisamos de un ajuste de regresión para realizar la estimación.

III. RESULTADOS EXPERIMENTALES

En esta sección presentamos los resultados de comparación de los métodos de predicción citados anteriormente, sobre un conjunto de las cuatro señales RSSI de la Fig. 1 ($RSSI_1, RSSI_2, RSSI_3$ y $RSSI_4$). La primera de ellas se capturó, mediante la utilidad *tcpdump* en *Linux Ubuntu 12.04 LTS 64 bit*, en un laboratorio del Campus Universitario de Tafira. Las tres últimas fueron adquiridas usando el programa *WireShark*. La $RSSI_4$ en *Windows 7 64 bit* y en otro laboratorio del mismo Campus Universitario; las otras dos en *Kali Linux*, leyendo balizas de dos encaminadores diferentes en una habitación de un piso, pero rodeado de AP de los vecinos, que en varios casos provocaban interferencias en el canal WiFi. Todas incluían información de TS de las balizas. Su número de medidas de RSSI es, respectivamente: 14958, 3733, 1946, y 3031. Como se puede apreciar se trata de señales tomadas en distintos entornos de uso real y frecuente por los usuarios de WiFi. Además, deliberadamente, se usaron diferentes sistemas operativos. Estas situaciones heterogéneas, tanto de software como de entornos geográficos, proporcionan una respuesta en entornos realistas, lo que pone de manifiesto la verdadera utilidad del método en situaciones cotidianas.

En la tabla I se muestra el *Error Cuadrático Medio (ECM)* entre cada una de las señales de trabajo y su predicción con cada uno de los métodos: modelo de *Ornstein-Uhlenbeck Link Prediction and Triggering (OU-LPT)* para el que hemos usado nuestra versión determinista implementada en [3] (en vez de la versión original de Chin et al. [6]), nuestro nuevo modelo de estimación de medias móviles con ventanas de tamaño variable: *Thresholded Local Gradient (TLG)*, *Gradiente (GRAD)* [4], *Gradiente Mejorado (MG)* [5] y filtrado de Kalman (KALMAN, el filtrado de Kalman que también fue analizado en [4] y es un clásico con el que se suelen comparar todos los métodos). Es de esperar que, respecto a los métodos GRAD y MG, TLG produzcan mejores resultados, tal como analizamos en la sección II.B. En las tablas I y II se puede apreciar claramente que TLG produce menores errores de estimación que GRAD y MG, consecuencia directa de que utilizan un tamaño fijo de ventana evaluada en la Ec. 11 y Ec. 14, respectivamente. Y

queda por demostrar si produce mejores resultados que KALMAN y OU-LPT.

Observamos que, en todos los casos, nuestro método es el que mejores resultados proporciona (el ECM es mínimo). Tanto GRAD como MG se quedan muy atrás en la comparativa, demostrando que no aprovechan correctamente la volatilidad de los X_k . Es de destacar el resultado de MG que, aunque segmenta con ventanas los X_k , no explota correctamente su evolución local. Esto se aprecia especialmente con la señal $RSSI_2$.

El método con resultados más cercanos al nuestro es OU-LPT. Pero está lejos de los excelentes resultados de TLG, considerando además las costosas operaciones estadísticas que precisa dicho método.

Por último, comprobamos que KALMAN es casi siempre el que más se aleja del mejor resultado.

En la tabla II se muestra los resultados de aplicar como parámetro de comparación la *Desviación Absoluta de la Media (DAM)*. La clasificación de resultados con respecto a aplicar ECM no cambia. Por tanto, podemos aseverar que TLG es un método novedoso, que muestra resultados mucho mejores que los otros métodos existentes para trazas de X_k heterogéneas y tomadas en escenarios cotidianos, lo que demuestra su utilidad práctica.

En las Fig. 3 a 7 presentamos las gráficas de estimación de cada uno de los métodos (en el mismo orden de las tablas) siendo la línea roja la estimación y los puntos negros los X_k . Sólo presentamos resultados en las gráficas para $RSSI_4$ por falta de espacio. El objetivo es dejar más claro por qué fallan los otros métodos y, en cambio, el nuestro (TLG) es más preciso como anteriormente se mostró en las tablas I y II. Por ello, tomando sólo una de las gráficas, se puede observar el comportamiento de los distintos métodos, que también se da en las otras señales. En la Fig. 4 se añade además asteriscos azules, que representan las muestras de las WT. Como las señales tienen demasiadas muestras, para apreciar los detalles, se ha seleccionado un único intervalo reducido de muestras en la gráfica de aplicación de cada método.

En la Fig. 3 se muestra que OU-LPT trata de seguir la tendencia de retorno a las medias móviles de RSSI a lo largo del tiempo. Pero, al considerar ventanas de tamaño fijo, no lo hace con la misma precisión en todos los tramos. Como consecuencia, cuando se producen saltos pronunciados (véase los instantes 154s y 164s), este método predice una trayectoria interpolada de RSSI que falsea el resultado.

Tabla I
COMPARACIÓN MEDIANTE ERROR CUADRÁTICO MEDIO

Señal	OU-LPT	TLG	GRAD	MG	KALMAN
$RSSI_1$	6.24734	0.03080	16.16566	9.86961	52.60055
$RSSI_2$	10.87396	0.03543	16.05904	741.96846	20.15416
$RSSI_3$	0.95903	0.02688	2.30898	1.37551	19.21611
$RSSI_4$	4.04108	0.01845	4.21207	4.61512	8.79967

Tabla II
COMPARACIÓN MEDIANTE DESVIACIÓN ABSOLUTA DE LA MEDIA

Señal	OU-LPT	TLG	GRAD	MG	KALMAN
$RSSI_1$	1.34196	0.12670	3.02587	1.42090	5.77475
$RSSI_2$	1.89079	0.12937	2.41630	3.86049	2.95100
$RSSI_3$	0.57254	0.09122	1.14523	0.57505	3.84875
$RSSI_4$	1.38355	0.05749	1.53697	1.29016	2.59320

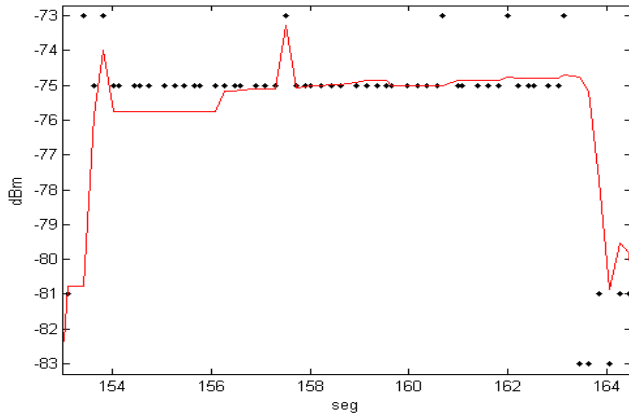


Fig. 3. Estimación bajo método OU-LPT.

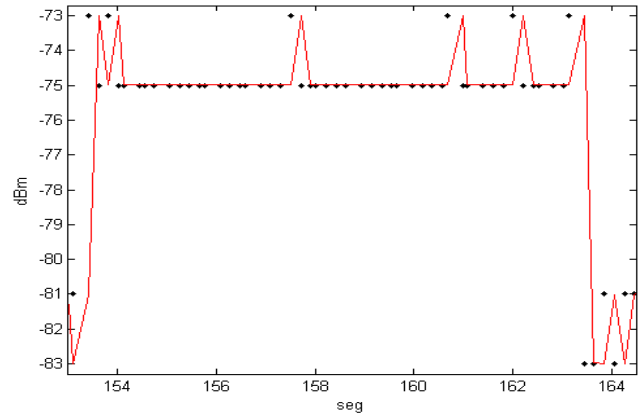


Fig. 6. Estimación bajo método MG.

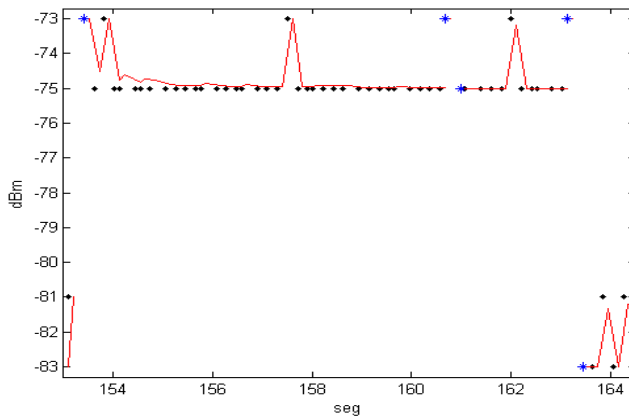


Fig. 4. Estimación bajo método TLG.

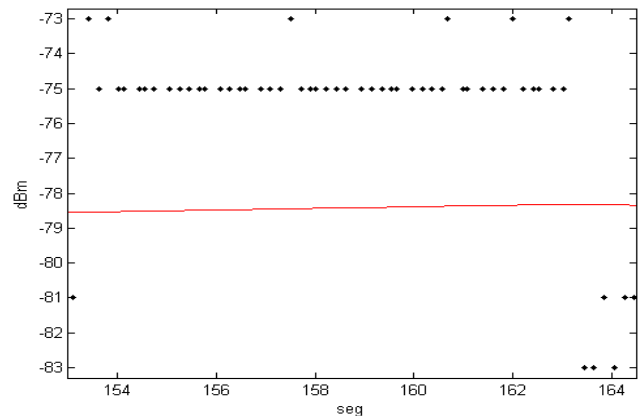


Fig. 7. Estimación bajo método KALMAN.

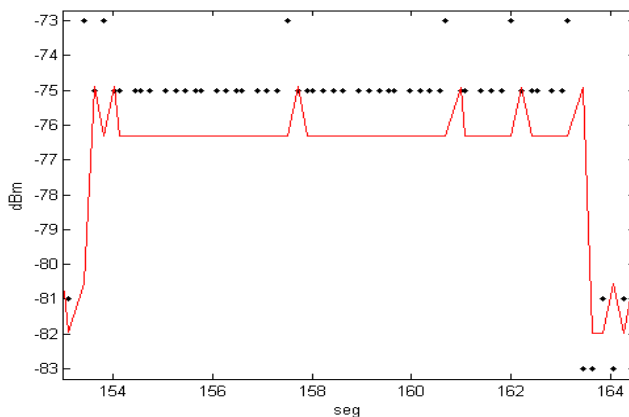


Fig. 5. Estimación bajo método GRAD.

La Fig. 4, por el contrario, ilustra cómo TLG trata de aproximarse el máximo a cada X_k . Recordemos que se ha fijado un umbral de amplitud de 0.5, porque el orden de cuantificación de la potencia real no nos asegura mayor precisión que ésa. Los asteriscos azules marcan el comienzo de una nueva ventana de medidas de RSSI, originada por saltos bruscos (instantes 154s y 164s) o por un retraso excesivo en la llegada de la siguiente baliza (p.e. en el instante 161s). Como observamos, nuestro método de construcción de ventanas adaptativas hace que la predicción se ajuste más a las X_k .

En la Fig. 5, correspondiente al método de GRAD, apreciamos que la falta de localidad a la hora de promediar las muestras provoca que toda la estimación se aleje de los puntos originales, por un *offset* considerable.

La predicción de GM se muestra en la Fig. 6. En este caso la señal se aproxima más a los puntos de referencia negros que en la Fig. 5. Pero al tomar ventanas de tamaño fijo la precisión es menor que en nuestro método, tal y como se refleja en las tablas I y II. También en este método se interpola la tendencia de la señal de potencia cuando se origina un salto brusco, lo que contribuye a falsear el resultado.

El último de los métodos se muestra en la Fig. 7, en la que se puede apreciar que KALMAN produce un suavizado excesivo sobre la señal, alejándose bastante de todos los puntos de referencia originales.

Como no es posible incorporar los resultados de todos los métodos para las señales utilizadas, en la Fig. 8 ilustramos solamente la estimación generada por nuestro algoritmo. De nuevo los puntos negros representan los X_k de las señales $RSSI_1$ a $RSSI_4$, siendo la línea roja los XP_k . Hemos suprimido las marcas de las WT, que con la resolución aplicada no serían visibles. A pesar de no mostrar explícitamente las WT, es fácilmente apreciable la partición de ventanas que ha efectuado nuestro método, especialmente las generadas por LUI sobre la señal $RSSI_1$. También hay que destacar el extenso LUI que se detecta en $RSSI_2$. Y, por último, reseñar las ventanas generadas por variación de amplitud en $RSSI_3$ y $RSSI_4$.

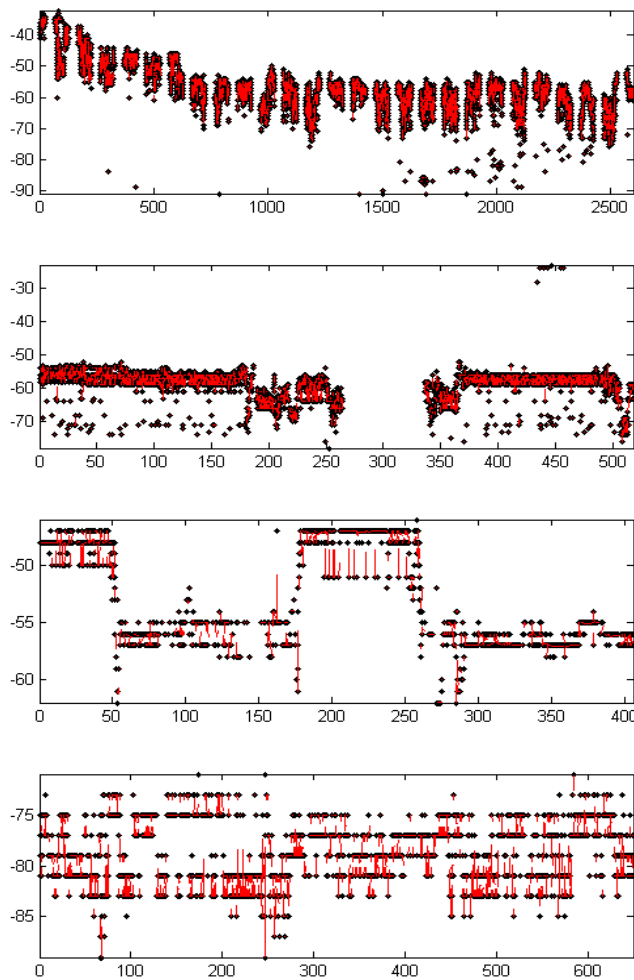


Fig. 8. Estimación TLG en $RSSI_t$ a $RSSI_s$ (de arriba a abajo).

En general el TLG muestra mejores resultados debido a que, ajustando el comienzo y tamaño de las ventanas para excluir los LUI y adecuarse al error de estimación, se logra una estimación más aproximada de las medias móviles a lo largo del tiempo. La adaptación del umbral del LUI a la variación local del retraso de balizas permite una prolongación de la ventana, lográndose estimar más X_k manteniendo la precisión del ajuste de la curva en los periodos sin información de balizas. La discriminación de los LUI es una medida prudente, puesto que éstos pueden haberse generado por incidencias en el canal. En ese caso el RSSI debe considerarse inexistente, y no tratar de estimarla como hacen otros métodos. El aislamiento en ventanas independientes de medidas de RSSI que supongan una variación brusca entre la ventana anterior y siguiente constituye un mecanismo flexible, que deja a la aplicación concreta el tratamiento de estas medidas desviadas. Un umbral de media unidad de cuantificación del RSSI (0.5 dBm) es suficiente para garantizar la precisión de la XP_k .

IV. APLICABILIDAD DEL MÉTODO

Nuestro método tiene una baja complejidad algorítmica, la estimación es más precisa que en el resto de métodos que conocemos, y no estima RSSI desacertadamente (en los LUI). Por ello, se puede pasar a los niveles superiores de la arquitectura de red valores más precisos y cualitativamente

mejores. No conocemos ningún método que tenga estas características.

Dado que la estimación del RSSI es un elemento clave en la definición de servicios inalámbricos eficientes en las redes actuales y futuras, analizamos brevemente cómo nuestro método puede impactar en la mejora de algunos de esos servicios empleando. El objetivo es exponer cómo puede influir en el diseño de servicios, que representen nuevos avances en servicios telemáticos.

En las redes inalámbricas de quinta generación (5G) se combina varias tecnologías de acceso radio. Una de éstas es WiFi, que debe inter-operar con otras tecnologías (redes heterogéneas). Para manejar la complejidad de esta interoperación, se debe incluir procedimientos de red inteligentes, que manejen adecuadamente niveles de interferencia, movilidad, información sobre el tráfico, entre otros [12]. Software *Defined Networking (SDN)* y *Network Function Virtualization* son dos tecnologías de diseño de arquitectura de redes muy actuales que permiten manejar esta complejidad de interoperación. Estas dos tecnologías necesitan de la abstracción de procedimientos a nivel físico y de enlace para proveer primitivas de diseño de servicios virtualizados en los niveles superiores de la arquitectura de red. Con nuestro método se puede proporcionar un nivel de abstracción que cuantitativa y cualitativamente proporcione información de mayor calidad que otros métodos. Por ejemplo, una *Application Programming Interface (API)* de SDN podría incluir el porcentaje de LUI en los últimos 5 segundos para informar acerca de las interrupciones intermitentes de X_k . Esto no lo podrían proporcionar otros métodos, porque siempre interpolan el X_k en las LUI.

Esta API se podría aplicar a la asociación eficiente de terminales inalámbricos a AP y a técnicas de handover.

Nuestro método puede guiar la asociación de terminales a AP [13] controlando el ancho de banda compartido por todos los terminales, proporcionando información sobre el BD; ya que, aunque desde el AP se reciba un RSSI elevado, puede que su respuesta sea demasiado lenta por estar gestionando demasiadas conexiones. Esto puede ser fácilmente controlado por nuestro método, porque puede recabar indirectamente (y pasarlo a los niveles superiores de la arquitectura de red mediante la API SDN) el estado del ancho de banda que puede suministrar el AP. La importancia de esto es clave, porque influye en el modelo de negocio de proyectos muy actuales como [2], que podrían redefinir el modelo de negocio de la conexión móvil a Internet en los próximos años.

El manejo eficiente del handover para procesos de *roaming* horizontal y vertical en las redes heterogéneas de 5G, usando técnicas de SDN, es un tópico importante actualmente [11,14,15]. Nuestra API podría proporcionar información de la estadística de LUI ocurridos recientemente, para que el nivel de aplicación pudiera estimarlos con elevada precisión, e iniciar el procedimiento de handover con más precisión.

En exteriores, la localización de terminales [16] es una de las tecnologías clave en la construcción de ciudades inteligentes (*Smart Cities*) usando 5G. En las redes vehiculares es fundamental la localización de terminales móviles [17] y en redes de sensores [18] es crucial el procedimiento de localización. Los sistemas de localización en interiores permiten definir servicios dependientes de la posición. Recientemente, en [19], se presenta un método que

utiliza el *Round Trip Time* y el *Signal Noise Ratio* de un solo AP (con varias antenas) para posicionar a un terminal inalámbrico comercial típico, sin necesidad de hardware adicional. Uno de los inconvenientes de estos sistemas de localización son los ciclos de histéresis, que se producen debido a la volatilidad del RSSI a lo largo del tiempo. Nuestro método evita ciclos de histéresis de predicción en periodos cortos de tiempo, porque busca la tendencia al retorno a las medias móviles del RSSI en el menor tiempo posible (del orden de un BI). Una API SDN que proporcione esta información ayudaría al nivel de aplicación a localizar mejor a los terminales.

En el encaminamiento en redes ad hoc inalámbricas [20], nuestro método ayudaría a no considerar que un terminal está conectado durante un LUI cuando realmente no lo está.

En sensado móvil participativo [21], al predecir con mayor precisión y tendencia a las medias móviles, se evitaría producir resultados en los que los participantes en el sensado serían detectados intermitentemente.

La última aplicación que citamos es la mitigación proactiva de los efectos de interrupciones del servicio de streaming móvil. En [22] se puede encontrar nuestro trabajo seminal que dio lugar a una línea extensa de trabajos posteriores, y recientemente se plantean trabajos de otros autores en esta línea [23]. Nuestro método puede ayudar a proactivamente informar sobre las interrupciones de servicio multimedia con la estadística de LUI, de las medias móviles recientes y los BD, para que la aplicación pueda anticipar proactivamente qué tipo de interrupciones podría ocurrir en el futuro próximo.

V. CONCLUSIONES

En este artículo hemos presentado un nuevo método de estimación de RSSI, que es capaz de detectar las medias móviles a las que retorna a lo largo del tiempo, definiendo una forma novedosa de construir las ventanas de medidas de RSSI. El algoritmo de estimación es de muy baja complejidad y eficiente, mejorando la estimación cualitativamente (en periodos de ausencia de RSSI en el terminal) y cuantitativamente (proporcionando un error reducido). Otra ventaja interesante es que, debido a la baja complejidad de nuestro algoritmo, se hace especialmente adecuado para estimar simultáneamente el RSSI de dos o más AP en procesos de *handover*, a diferencia de otros cuya complejidad algorítmica es mucho más elevada.

Este método, además, abre líneas interesantes. Entre ellas: la mejora del LUI, estimar el RSSI durante los LUI aplicando otro algoritmo que proporcione una precisión o eficiencia más reducida (si la aplicación lo requiere), y optimizar los algoritmos de *handover* considerando LUI consecutivos, combinados o no con variaciones bruscas del RSSI.

AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación parcial de la ULPGC.

REFERENCIAS

[1] K. Atalah, "Contributions to effective protocol design to mitigate continuous multimedia services short disruptions in WiFi networks, Ph.D. Thesis (A. Suarez & Elsa Macias supervisors)," 2011.
 [2] Google, "Project Fi," [Online; accessed 28-April-2015] <<https://fi.google.com/about/network/>>, 2015.

[3] A. Suárez, J.A. Santana, E.M. Macias-Lopez, V.E. Mena, J.M. Canino, and D. Marrero, "RSSI Prediction in WiFi Considering Realistic Heterogeneous Restrictions," *Network Protocols and Algorithms*, vol. 6, 2014, pp. 19–40.
 [4] A. Suárez, K.A. Elbatsh, and E. Mactas, "Gradient rssi filter and predictor for wireless network algorithms and protocols," *Network Protocols and Algorithms*, vol. 2, 2010, pp. 1–26.
 [5] F. Subhan, S. Ahmed, K. Ashraf, and M. Imran, "Extended Gradient RSSI Predictor and Filter for Signal Prediction and Filtering in Communication Holes," *Wireless Personal Communications*, 2015, pp. 1–18.
 [6] E. Chin, D. Chieng, V. Teh, M. Natkaniec, K. Loziak, and J. Gozdecki, "Wireless link prediction and triggering using modified Ornstein–Uhlenbeck jump diffusion process," *Wireless Networks*, vol. 20, 2014, pp. 379–396.
 [7] G.K. Smith, "Beacon Jitter Prediction for Wireless Local Area Network (LAN) Devices," Google Patents, 2014.
 [8] M. van der Schaar and S.S. N, "Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms," *Wireless Communications, IEEE*, vol. 12, Aug. 2005, pp. 50–58.
 [9] N.A. Ali, E. Ekram, A. Eljasmy, and K. Shuaib, "Measured delay distribution in a wireless mesh network test-bed," *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on, IEEE*, 2008, pp. 236–240.
 [10] X. Long and B. Sikdar, "A real-time algorithm for long range signal strength prediction in wireless networks," *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, IEEE, 2008, pp. 1120–1125.
 [11] A.S. Sadiq, K.A. Bakar, K.Z. Ghafoor, J. Lloret, and S. Mirjalili, "A smart handover prediction system based on curve fitting model for Fast Mobile IPv6 in wireless networks," *International Journal of Communication Systems*, vol. 27, 2014, pp. 969–990.
 [12] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5G on the Horizon: Key Challenges for the Radio-Access Network," *Vehicular Technology Magazine, IEEE*, vol. 8, 2013, pp. 47–53.
 [13] D. Gupta, P. Mohapatra, and C.-N. Chuah, "Seeker: A bandwidth-based association control framework for wireless mesh networks," *Wireless Networks*, vol. 17, 2011, pp. 1287–1304.
 [14] J. Geurts, Z. Li, Y. Liu, and J.-C. Point, "Transparent Handover Using WiFi Network Prediction for Mobile Video Streaming," *New Contributions in Information Systems and Technologies*, A. Rocha, A.M. Correia, S. Costanzo, and L.P. Reis, eds., Springer International Publishing, 2015, pp. 937–946.
 [15] M.F. Tuysuz and H.A. Mantar, "Minimizing Communication Interruptions Using Smart Proactive Channel Scanning Over IEEE 802.11 WLANs," *Wireless Personal Communications*, 2015, pp. 1–26.
 [16] S. Chatterjee, P. Banerjee, and M. Nasipuri, "Enhancing Accuracy of Localization for Primary Users in Cognitive Radio Networks," *Computational Intelligence and Networks (CINE), 2015 International Conference on*, 2015, pp. 74–79.
 [17] J. Liu, J. Wan, Q. Wang, P. Deng, K. Zhou, and Y. Qiao, "A survey on position-based routing for vehicular ad hoc networks," *Telecommunication Systems*, 2015, pp. 1–16.
 [18] B. Mukhopadhyay, S. Sarangi, and S. Kar, "Performance evaluation of localization techniques in wireless sensor networks using RSSI and LQI," *Communications (NCC), 2015 Twenty First National Conference on*, 2015, pp. 1–6.
 [19] C. Yang and H. Shao, "WiFi-based indoor positioning," *Communications Magazine, IEEE*, vol. 53, Mar. 2015, pp. 150–157.
 [20] L. Hong, X. Liu, L. Zhang, and W. Chen, "Towards sensitive link quality prediction in ad hoc routing protocol based on grey theory," *Wireless Networks*, 2015, pp. 1–11.
 [21] L. Wang, W. Liu, N. Jing, and X. Mao, "Simultaneous navigation and pathway mapping with participating sensing," *Wireless Networks*, 2015, pp. 1–19.
 [22] Á. Suárez, M. La-Menza, E.M. Macías, and V.S. Sunderam, "Automatic Resumption of Streaming Sessions over Wireless Communications Using Agents," *IMECS*, 2006, pp. 926–931.
 [23] V. Siris, M. Anagnostopoulou, and D. Dimopoulos, "Improving Mobile Video Streaming with Mobility Prediction and Prefetching in Integrated Cellular-WiFi Networks," *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, I. Stojmenovic, Z. Cheng, and S. Guo, eds., Springer International Publishing, 2014, pp. 699–704.

Análisis de las Prestaciones de un Esquema de Codificación Lineal Aleatoria sobre Redes Inalámbricas

Pablo Garrido[†], David Gómez[†], Jorge Lanza[†], Ramón Agüero[†] y Joan Serrat[¶]

[†]Universidad de Cantabria

Avda. de los Castros s/n.

39005 Santander, España

{pgarrido,dgomez,ramon,jlanza}@tmat.unican.es

[¶]Universitat Politècnica de Catalunya

C/ Girona 31

08034 Barcelona, España

serrat@tsc.upc.edu

Resumen—Este trabajo presenta un novedoso esquema de *Network coding intra-flujo*, que une un esquema de codificación aleatoria (*Random Linear Combination, RLC*) con el protocolo *UDP*. Esta combinación ofrece un servicio fiable, que garantiza la recepción ordenada de todos los mensajes enviados por el nivel de aplicación de la estación transmisora. Así, se compara sus prestaciones con las de una transmisión *TCP*, solución tradicional para ofrecer el mismo tipo de servicio, a través de una extensa campaña de simulación sobre redes *IEEE 802.11*, con la herramienta *ns-3*. Los resultados ponen de manifiesto una mejora de hasta un 20% sobre un canal ideal. Se propone además el intercambio de una semilla aleatoria, para aliviar la sobrecarga de las cabeceras del protocolo. Por otro lado, cuando las condiciones del canal empeoran, la solución conseguirá enmascarar la aparición de errores, manteniendo un *throughput* notablemente superior al ofrecido por *TCP*.

Palabras Clave—Redes inalámbricas, Network Coding, Codificación Lineal Aleatoria, TCP, Análisis de rendimiento

I. INTRODUCCIÓN

En los últimos años se ha observado un claro cambio de tendencia en lo que se refiere a las conexiones denominadas *de última milla*, de tal manera que las tecnologías inalámbricas se están imponiendo a las alternativas cableadas. La proliferación de los dispositivos móviles (por ejemplo *smartphones*, *tablets*, y un largo etcétera) proporcionan al usuario un gran abanico de posibilidades a la hora de acceder a la red.

Uno de los principales problemas que aparecen es que *TCP*, el protocolo de nivel de transporte más utilizado en la actualidad, presenta un comportamiento deficiente sobre este tipo de enlaces [1]. Los mecanismos de control de congestión que incorpora, (mal)interpretan toda pérdida de segmentos como producto de la congestión en los nodos intermedios, a pesar de que las pérdidas se deben mayoritariamente a la deficiente calidad del medio inalámbrico. La consecuencia más inmediata es un deterioro notable en el rendimiento de la transmisión.

Una de las soluciones más prometedoras para paliar este problema son las técnicas conocidas como *Network Coding (NC)*, que proponen un cambio radical del concepto más clásico de encaminamiento. Mientras que las comunicaciones tradicionales se basan en un esquema "*store-and-forward*", según el que los nodos intermedios se limitan a reenviar los paquetes que reciben hacia su siguiente salto, el paradigma de *NC* es el "*compute-and-forward*". Entre varias

de las ventajas que aportan estas técnicas, este trabajo se centra en el aumento del pobre rendimiento ofrecido por *TCP* sobre redes inalámbricas.

Este trabajo parte de uno de los enfoques que ha suscitado más interés de *NC*, basado en la codificación de paquetes pertenecientes a un mismo flujo, y que se conoce como esquema *intra-flujo*¹. Mediante la utilización de un esquema de codificación aleatoria o *Random Linear Coding (RLC)*, llevado a cabo en primera instancia por el origen, se transmitirán combinaciones lineales de bloques formados por un número constante de datagramas. Con esto, será suficiente con que el destino reciba un número determinado de paquetes codificados linealmente independientes entre sí, para recuperar el contenido generado por la aplicación transmisora, sin que sea relevante el orden de llegada.

El documento se ha dividido como se recoge seguidamente: en la Sección II se resumen las contribuciones más interesantes de la literatura relacionado con el empleo de *NC* en entornos inalámbricos. La Sección III describe la implementación del protocolo *NC* llevada a cabo en este trabajo. La Sección IV muestra la configuración utilizada para llevar a cabo las simulaciones que se utilizarán para evaluar el rendimiento de la solución presentada, cuyos resultados más interesantes se resumen en la Sección V. Por último, la Sección VI concluye el documento y resume aquellos puntos que se acometerán en el futuro.

II. TRABAJOS RELACIONADOS

El concepto de *Network Coding* aparece inicialmente en el año 2000, cuando Ahlswede *et al.* [4] proponen una aproximación novedosa para el encaminamiento de paquetes. Tradicionalmente, los nodos intermedios o *routers* tenían la tarea *única* de reenviar los datagramas recibidos hacia un siguiente salto que encontraban en sus tablas de encaminamiento, indexadas por la dirección del destino final. Sin embargo, *NC* otorga a estos nodos una capa de inteligencia adicional, que les permitirá procesar y modificar el contenido de los paquetes a medida que los atraviesan.

¹La otra opción, conocida como *inter-flujo* o *inter-sesión*, basa su operación en mezclar el tráfico de diferentes flujos en los nodos intermedios, y no forma parte del análisis realizado en este trabajo. Otros trabajos [2], [3] han mostrado las dificultades de este tipo de esquemas

A partir de esta propuesta inicial, y tras una primera etapa con un carácter más teórico [5], se demostró que el uso de combinaciones lineales (operaciones con cuerpos finitos [6]) permitía alcanzar las cotas superiores de rendimiento, sin recurrir a complejos esquemas de codificación; posteriormente se demostró que se podía incluso elegir los coeficientes de manera aleatoria [7], sin perjudicar el rendimiento.

Ha habido un número de trabajos que han propuesto soluciones basadas en las técnicas de *NC* para mejorar las comunicaciones sobre redes inalámbricas. Se puede destacar en primer lugar el protocolo *COPE* [8], en el que los nodos intermedios utilizan operaciones *XOR* para “mezclar” el tráfico perteneciente a diferentes flujos, consiguiendo reducir el número de transmisiones necesarias para que la comunicación finalice. No obstante, los autores observan que, al emplear *TCP*, las pérdidas originadas por el efecto del terminal oculto deterioran notablemente el rendimiento, debido a los mecanismos de control de congestión utilizados por *TCP*. Además, el uso de este tipo de esquemas de codificación *inter-flujo* lleva asociado una elevada sobrecarga, fruto de los mensajes de señalización necesarios.

Con una aproximación diferente, Chachulski *et al.* presentan en [9] el protocolo *MAC independent Opportunistic Routing (MORE)*, una solución que combina los paquetes pertenecientes a una misma conexión, guardando ciertas similitudes con los códigos *Digital Fountain*, como *Luby Transform (LT)* [10] o *Rapid Tornado (RAPTOR)* [11], aunque el uso de *NC* permite que los nodos intermedios puedan manipular y transformar la información. A través de un esquema de codificación aleatoria de bloques de tamaño fijo, los autores se centran en la caracterización de un protocolo de encaminamiento oportunista sobre despliegues multi-salto. Sin embargo, no se centran en analizar la relación de su propuesta (situada entre las capas 2 y 3) con protocolos de nivel de transporte, como *TCP*, por lo que no se analiza si el retardo asociado a la transmisión de bloques de mayor tamaño podría incrementar de manera significativa los valores del *Round Trip Time (RTT)* estimado en la fuente, lo que perjudicaría los mecanismos de control de congestión de *Transport Control Protocol (TCP)*. Otra solución que ha acaparado gran interés es la de Sundararajan *et al.* [12], que proponen mejorar el comportamiento de *TCP*, a través de un protocolo de codificación aleatoria que sitúan entre las capas de transporte y de red. Aprovechando los mecanismos de control de flujo y congestión, su principal contribución es un sistema de reconocimientos propietarios, en el que el destino confirma la recepción de información útil. Gracias a estas confirmaciones, los autores proponen un mecanismo de ventana deslizante en el que los mensajes originales se recuperan dinámicamente. Se consiguen notables mejoras en el rendimiento frente al ofrecido por la versión tradicional de *TCP* sobre canales con errores, aunque el número de mensajes de señalización podría incrementarse de manera notable, lo que tendría una consecuencia negativa en la contención del canal.

Las principales limitaciones que se pueden achacar a los esquemas de codificación *intra-flujo* son las siguientes: (1) la sobrecarga asociada a la transmisión de los coeficientes

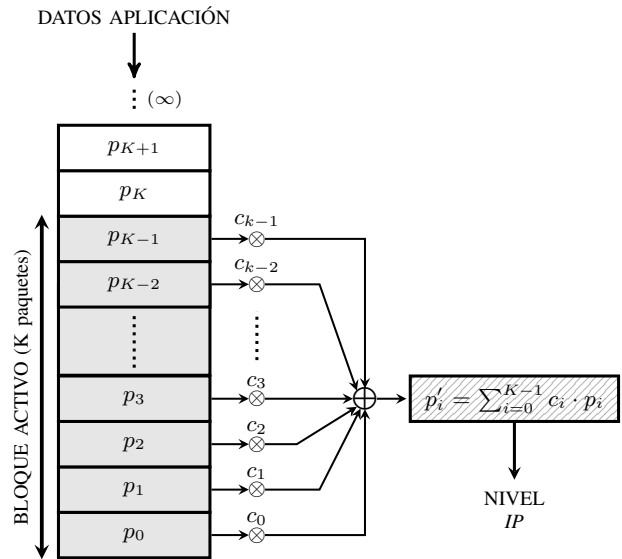


Fig 1: Codificación en el nodo fuente

utilizados para combinar el contenido de los mensajes; (2) el coste computacional asociado a las tareas de codificación, recodificación y decodificación. Existen trabajos que buscan limitar el efecto de estos problemas, como el citado protocolo *CTCP* [13], que evita la transmisión de los coeficientes gracias al uso de un generador de números pseudo-aleatorios y la transmisión de la semilla correspondiente, u otras alternativas como [14] o [15], que utilizan patrones de codificación para reducir la complejidad de las operaciones en las que se basan estos esquemas.

El trabajo presentado en este artículo tiene como principal objetivo proponer una solución que ofrezca un servicio equivalente al de *TCP*, garantizando por tanto la entrega ordenada de toda la información enviada por la fuente; para ello se combinan un esquema de codificación *NC intra-flujo*, y el protocolo *UDP*. Además, se introducen diferentes técnicas *cross-layer*, para optimizar la inyección de tráfico hacia los niveles inferiores, evitando además la transmisión de información redundante.

III. PROTOCOLO INTRA-FLUJO

Esta sección describe el proceso seguido para el intercambio de información entre los nodos origen y destino. Como se ha adelantado, se utiliza para ello la combinación del protocolo *UDP* con una solución de codificación aleatoria o *RLC*, que se sitúa como una entidad independiente entre las capas de red y transporte, haciendo además las veces de protocolo *NC*, adaptado para aprovechar las propiedades del medio inalámbrico. Su implementación se ha llevado a cabo en la plataforma de simulación *ns-3* [16].

Tal y como representa la Fig. 1, el primer elemento clave es la operación de codificación, realizada inicialmente por el nodo fuente². El nivel *NC* incorpora un *buffer* de capacidad ilimitada para almacenar los paquetes generados por la aplicación. Teniendo en cuenta que el parámetro *K* establece el

²Los nodos intermedios tienen la capacidad de modificar el contenido de los mensajes codificados. Sin embargo, este trabajo únicamente analiza el comportamiento extremo a extremo de la transmisión.

número de paquetes que componen un bloque, una vez que el *buffer* tenga almacenados un mínimo de K datagramas de datos, la entidad *NC* transmisora comenzará la fase de codificación. Así, y como puede apreciarse en la figura, se enviará a los niveles inferiores un paquete codificado p'_j , generado a través de la combinación lineal de los K paquetes p_i , $i \in [1, K]$, utilizando un conjunto de coeficientes c_i , elegidos aleatoriamente entre los valores de cuerpo finito de orden $GF(2^q)$. Estos K elementos se representan en forma de vector, \vec{c}_i , al que se denomina “vector de coeficientes”. La Ec. 1 resume la operación descrita para generar un paquete codificado p'_j .

$$p'_j = \sum_{i=0}^{K-1} c_i \cdot p_i \quad (1)$$

Una propiedad clave de este tipo de esquemas es su característica “rateless”, según la cual la fuente puede generar, de manera indefinida, combinaciones lineales potencialmente diferentes entre sí, hasta que el destino consiga recuperar la información original. De este modo, el transmisor continuará enviando paquetes codificados de manera continuada, hasta que reciba un mensaje de confirmación por parte del receptor, cuando procederá a eliminar los K paquetes que componen el bloque ya decodificado. Como consecuencia, se puede decir que los paquetes codificados carecen de significado de manera aislada, siendo suficiente con que el receptor reciba una cantidad mínima de paquetes para decodificar el bloque completo, sin importar el orden o la identidad de los mensajes recibidos. De esto se deduce que, en este esquema de codificación en particular cada mensaje codificado transporta una fracción equivalente a $\frac{1}{K}$ del total del bloque. Además, otra de las ventajas asociadas a este tipo de transmisión “continua” es que consigue enmascarar la presencia de errores en el canal, reduciendo su impacto en el rendimiento.

Por su parte, el nodo receptor recoge los paquetes codificados y procesa su contenido. Para ello, se implementan dos contenedores diferentes: un *buffer de decodificación*, capaz de guardar hasta K paquetes codificados, que serán utilizados posteriormente para recuperar el bloque original. Además, hace falta una matriz de dimensión $K \times K$, inicialmente vacía, donde se almacenan y procesan los vectores de coeficientes \vec{c}_i . En el protocolo se han considerado dos alternativas para transportar dichos vectores: de manera explícita, incluyendo todos los elementos en el propio vector; o siguiendo un esquema similar a la propuesta en [13], transmitiéndose una semilla que será utilizada posteriormente en recepción para reconstruir el vector de coeficientes original, haciendo uso de un generador de números pseudo-aleatorios. El procedimiento es el siguiente: la entidad *NC* mantiene una variable interna con el rango de la matriz de coeficientes; tras la recepción de un paquete codificado, se añade el vector de coeficientes \vec{c}_i a la matriz; posteriormente, se comprueba su rango: si ha aumentado, se considera que el vector recibido es linealmente independiente a los que ya se tenían almacenados, por lo que se incrementa el rango actual y se mantienen el vector, dentro de la matriz, y el paquete codificado, en el *buffer de decodificación*. En caso contrario (el vector es una combinación lineal

de los anteriores³), el paquete recibido quedará descartado, eliminando el vector de la matriz.

Este proceso se irá repitiendo hasta que la matriz de coeficientes tenga rango completo (K) y sea invertible, cuando se pasa a decodificar la información, tal y como recoge la Ec. 2. A partir de ese momento, se descartan todas las recepciones posteriores del mismo bloque generando el envío de un nuevo reconocimiento. Tras recuperar la información original, los K datagramas nativos, el receptor notificará a la fuente, a través de un mensaje de confirmación, la correcta decodificación, por lo que ésta podrá borrar su contenido, pasando a la transmisión del siguiente bloque. Así, un nodo destino necesitará recibir un total de K paquetes codificados con vectores linealmente independientes. En una transmisión, el valor real de paquetes necesarios será K' , siendo $K' \geq K$, ya que: (1) pueden enviarse vectores que son combinación lineal de los anteriores y (2) las condiciones del canal pueden hacer que se pierdan mensajes durante la transmisión.

$$\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{K-1} \end{bmatrix} = \begin{bmatrix} c_{0,0} & \cdots & c_{0,K-1} \\ c_{1,0} & \cdots & c_{1,K-1} \\ c_{2,0} & \cdots & c_{2,K-1} \\ \vdots & \ddots & \vdots \\ c_{K-1,0} & \cdots & c_{K-1,K-1} \end{bmatrix}^{-1} \cdot \begin{bmatrix} p'_0 \\ p'_1 \\ p'_2 \\ \vdots \\ p'_{K-1} \end{bmatrix} \quad (2)$$

A modo de ejemplo, en la Fig. 2 se ilustran los eventos producidos en los nodos origen y destino durante la transmisión de dos bloques consecutivos, tanto desde el punto de vista del transmisor (parte superior de la figura) como del receptor (parte inferior). Como se aprecia en el diagrama correspondiente a la codificación, se generan combinaciones lineales con los K paquetes que conforman el bloque activo. En el primero de ellos, se ha representado una situación en la que todos los paquetes codificados (p'_i) transportan información “linealmente independiente”, por lo que el rango de la matriz crece tras cada recepción, siendo éste el supuesto óptimo, $K' = K$. Una vez completada la decodificación, el receptor envía los K datagramas al nivel superior, y transmite la correspondiente confirmación a la fuente, que procede a eliminar el contenido del bloque, pasando a la transmisión del siguiente. En este segundo bloque, se observan varias circunstancias que harán crecer el número de eventos necesarios para completar correctamente el proceso de transmisión de un bloque: recepción de paquetes con vectores que son combinación lineal de los ya almacenados, o pérdidas producidas por el medio inalámbrico, tanto de paquetes codificados como de reconocimientos. Como era de esperar, hasta que la fuente no reciba correctamente un reconocimiento, seguirá transmitiendo paquetes correspondientes a un bloque que el receptor ya ha recuperado, lo que generará un uso no útil del canal, penalizando al rendimiento de la comunicación. Hay que tener en cuenta además que la transmisión de los reconocimientos genera una contención adicional en el acceso al canal.

Tras describir la operación del protocolo, se construye una cabecera propia del nivel *NC* que contiene los siguientes campos:

³Hay que tener en cuenta que el transmisor elimina la posibilidad de transmitir vectores de coeficientes nulos.

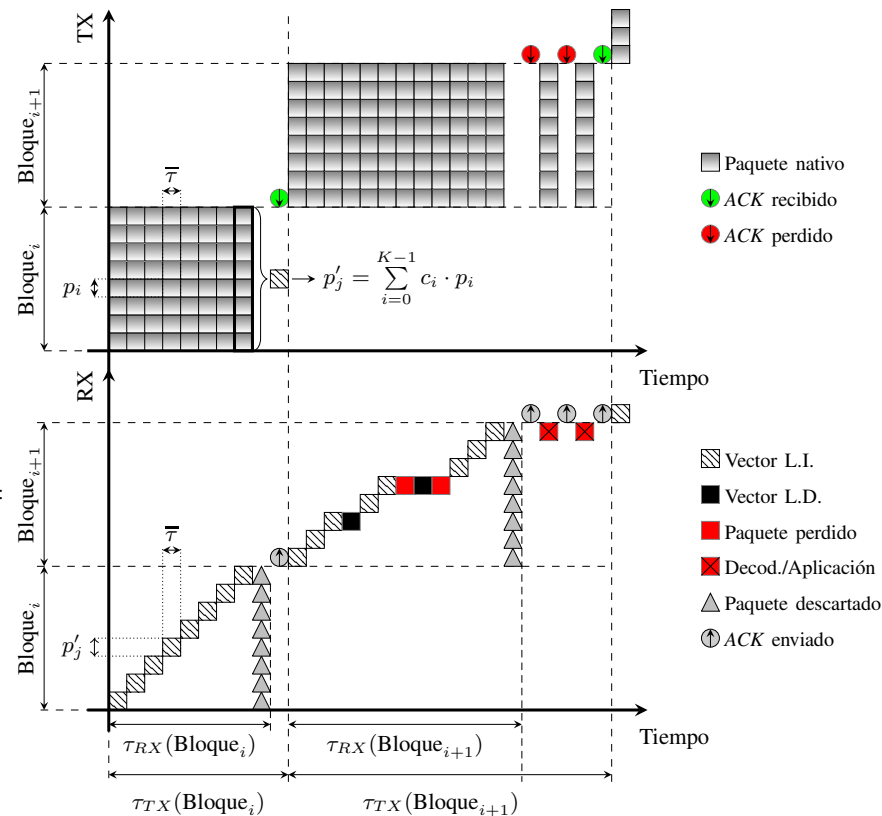


Fig 2: Secuencia de eventos discretos para una transmisión arbitraria

- **Tipo de mensaje** (1B). Este campo distingue, en la versión actual del protocolo, si se trata de un mensaje de datos o paquete codificado ('0') o, por el contrario, de un reconocimiento ('1').
- **Tamaño del bloque** - K (1B). Número de paquetes que conforman un bloque de información, limitado a 256.
- **Orden del cuerpo finito** - $GF(2^q)$ (1B). Exponente que actuará como extensión sobre el cuerpo binario. La librería utilizada para llevar a cabo las operaciones aritméticas sobre estos cuerpos limita el tamaño máximo a $q = 8$.
- **Puertos UDP origen y destino** (4B). La información correspondiente a la cabecera UDP viaja encriptada en el interior de un paquete codificado, por lo que se incluirán los números de puerto, tanto origen como destino, para identificar el flujo al que pertenece cada paquete.
- **Vector de coeficientes** (1 - 256B). Es necesario transmitir los K coeficientes aleatorios con los que se combinan los K paquetes del bloque. Se manejan dos posibilidades: en primer lugar, se transmite el vector de manera explícita, ocupando un tamaño que responde a la expresión $\lceil \frac{K \cdot q}{8} \rceil$ bytes, variando desde 1 octeto hasta los 256; por otra parte, si se utiliza la opción de enviar la semilla utilizada en el generador de números pseudoaleatorios, se consigue fijar la longitud a un valor de 4 octetos, independiente del tamaño del bloque o el cuerpo en el que se trabaje.

El paquete de reconocimiento, enviado por el destino a la fuente cuando este ha decodificado un bloque, también

contiene los campos de la cabecera, excepto el vector de coeficientes. Por lo que cada paquete de reconocimiento identifica el número de bloque que confirma, pudiendo ignorar la fuente reconocimientos duplicados.

Por último, se puede destacar la utilización de diferentes mecanismos *cross-layer*, utilizados para conectar la entidad *NC* con los niveles inferiores. En primer lugar, se ha habilitado el *modo promiscuo en los nodos*, de manera que la capa *NC* pueda procesar todos los paquetes transmitidos dentro de su región de cobertura; esto facilitará además la aparición de múltiples rutas entre el origen y destino, posibilitando la utilización de soluciones de encaminamiento oportunista. Por otra parte, la inyección de los paquetes codificados hacia los niveles inferiores en la fuente (funcionalidad similar a un control de flujo) se llevará a cabo a una frecuencia óptima, que adaptándose a la situación del medio físico; así, se enviará una señal cada vez que se transmita al canal un paquete codificado, de manera que el nivel *NC* sea consciente que puede proceder a enviar el siguiente, con lo que la tasa de transmisión vendrá dada por la propia capacidad del canal. Finalmente, se ha habilitado un mecanismo de borrado selectivo en el *buffer* del nivel de enlace, mediante el cual, tras la recepción de una confirmación, el nodo origen podrá eliminar exclusivamente aquellos mensajes de bloques obsoletos, evitando transmisiones innecesarias. Esto será especialmente útil en situaciones con varios flujos activos de manera simultánea.

IV. PLATAFORMA DE MEDIDAS

Para evaluar el comportamiento de la solución propuesta se ha llevado a cabo una exhaustiva campaña de simulaciones

Tabla I: Parámetros de la simulación

Parámetro	Valor
Nivel físico	IEEE 802.11b (11 Mbps)
Errores	Modelo nativo
TX IEEE 802.11	1, 4, 7
RTS/CTS	Deshabilitado
Tamaño del bloque (K)	2, 4, 8, 16, 32, 64, 128, 256
Orden del cuerpo GF(2 ^q)	q ∈ [1, 8]
Capa de transporte	UDP (NC)/TCP
Aplicación	$\lceil \frac{10000}{K} \rceil$ paquetes
Tasa binaria (aplicación)	Constante (Saturación)
Tráfico	Unicast
Payload	1500 bytes en nivel IP

sobre la plataforma ns-3 [16], comparando el rendimiento de diferentes alternativas sobre un enlace directo IEEE 802.11. La Tabla I presenta los parámetros que definen el escenario.

Se sigue el método Monte Carlo para asegurar la validez estadística de los resultados, es decir, se realizan 100 simulaciones por configuración y se obtienen los valores medios y los intervalos de confianza del 95 % para las diferentes métricas estudiadas.

En la Sección V se presentan los principales resultados, obtenidos sobre un enlace inalámbrico directo, punto a punto (origen/destino). Como ya se ha descrito anteriormente, en el esquema NC presentado en este trabajo, la fuente enviará paquetes codificados a la tasa a la que le permita el medio físico, en condiciones de saturación de la red. En una primera etapa se caracterizará el comportamiento del protocolo *intra-flujo* bajo condiciones ideales, para analizar la influencia de la configuración de codificación escogida. Posteriormente, se introducirán errores en el enlace entre los nodos, estudiando así la capacidad de la solución propuesta para enmascarar las pérdidas producidas en el canal. En ambos casos, se comparará el rendimiento con el observado en una transmisión TCP tradicional. Comentar finalmente que, en el caso de utilizar el protocolo NC, la capa de aplicación generará paquetes de longitud máxima (para una MTU de 1500 bytes), para lo que tendrá que descontar, a los típicos 1472 bytes del datagrama UDP, la longitud de la cabecera del protocolo NC.

V. RESULTADOS

En una primera etapa, se ha utilizado un enlace ideal entre los nodos origen-destino, modificando el tamaño del bloque y el orden del cuerpo finito sobre el que se definen las operaciones efectuadas sobre los datagramas nativos, K y q, respectivamente. Antes de estudiar el rendimiento de la transmisión, es importante analizar la influencia de los parámetros de configuración, K y q, sobre el número de transmisiones que deben realizarse para completar de manera exitosa la transmisión de un bloque. Teniendo en cuenta que se está transmitiendo sobre un canal sin errores, se puede asegurar, que el único factor que podría incrementar el número de transmisiones, K', con respecto al de una transmisión óptima, K, es la generación de vectores que son combinaciones lineales de los ya recibidos previamente por el destino, que tendría que descartarlos, por no aportar información novedosa.

La Fig. 3 recoge la función de distribución del número de paquetes en exceso por bloque, representando en el eje de abscisas la variable $\delta = K' - K$, en función de los

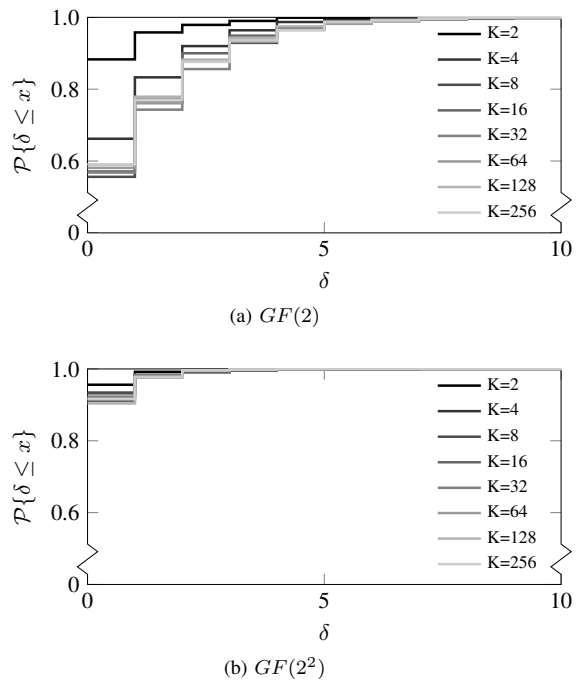


Fig 3: Función de distribución del número de paquetes en exceso

tamaños de las matrices de coeficientes (tamaño del bloque) y del orden del cuerpo finito, mostrando los resultados para el cuerpo binario GF(2) (Fig. 3a) y para su extensión más inmediata, GF(2²) (Fig. 3b). En ambos casos se observa que la probabilidad de que se produzca una transmisión ‘ideal’, esto es, $\delta = 0$, decrece a medida que aumenta el tamaño del bloque; no obstante, hay que tener en cuenta que, a medida que aumenta el valor de K, menor será la penalización debida a la transmisión de paquetes en exceso. En el estudio se puso de manifiesto asimismo una mejora notable al aumentar el orden del cuerpo finito, obteniéndose probabilidades $\mathcal{P}\{\delta \leq 0\} \approx 1$ para $q > 2$ (por motivos de espacio, estos resultados no se han representado en este artículo). Sin embargo, el uso de cuerpos de mayor orden también implicaría una mayor sobrecarga en la cabecera (cuando el vector de coeficientes se transmite de manera explícita). Finalmente, debería tenerse en cuenta el impacto adicional de los mensajes de reconocimiento transmitidos tras la decodificación exitosa de un bloque, lo que penalizará en mayor medida la comunicación, al utilizar valores de K bajos.

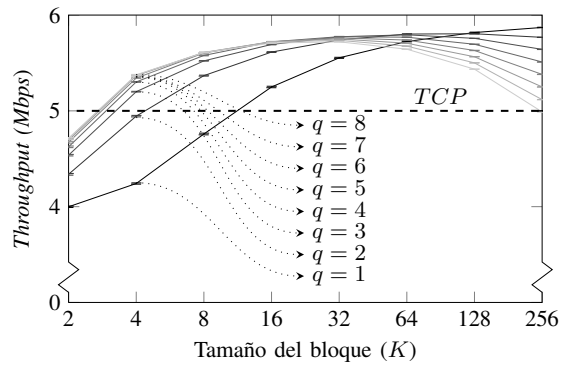
La Fig. 4 muestra el rendimiento de la comunicación para las dos versiones del protocolo que se han propuesto, en función del envío explícito o no del vector de coeficientes. En la versión más básica (Fig. 4a), los coeficientes utilizados para la generación del paquete codificado se añaden a la cabecera, por lo que, como se describió en la Sección III, su longitud variable estará entre 1 y 256 bytes, dependiendo de los parámetros K y q utilizados. Hay que tener en cuenta que el tamaño total de la cabecera NC, en bytes, se corresponde con la expresión $\lceil \frac{K \cdot q}{8} \rceil$. A medida que crezca el tamaño de los coeficientes del numerador, menor será el espacio disponible en la trama, perjudicando por tanto el rendimiento neto. En la figura se pueden destacar dos aspectos: en primer lugar,

cuando el tamaño del bloque es pequeño ($K \leq 64$), se ve que el rendimiento crece con K y, además, la utilización de cuerpos de mayor orden aporta una ganancia adicional. Sin embargo, cuando el tamaño del bloque es más elevado ($K > 64$), a pesar de que se ha comprobado que el uso de cuerpos de mayor orden permite reducir el número de transmisiones espúreas (paquetes que no aportan información, por ser combinación lineal de los ya recibidos), la sobrecarga introducida por los coeficientes tiene como consecuencia un empeoramiento del *throughput* a medida que q aumenta. Para este caso, la configuración óptima se obtiene para unos valores de $q = 1$ y $K = 256$.

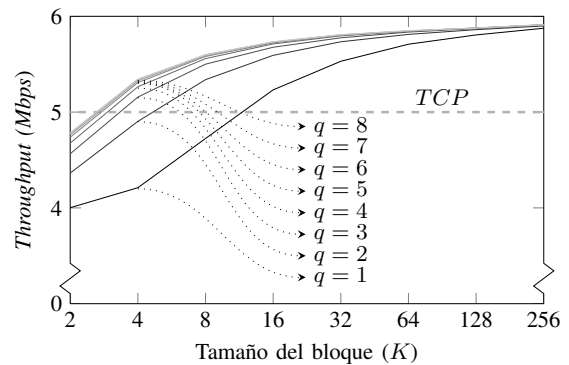
En una mejora del protocolo, la transmisión de la semilla aleatoria permite mantener un tamaño de cabecera fijo de 13 bytes, reduciendo significativamente la sobrecarga y permitiendo, por tanto, incrementar la cantidad de información que se puede incluir por paquete codificado. La Fig. 4b muestra los valores de *throughput* obtenidos con esta versión, observándose que es creciente tanto con el tamaño del bloque como con el orden del cuerpo, al contrario de lo que sucedía con la configuración previa. En relación al comportamiento de TCP, los mecanismos de control de congestión y la transmisión de reconocimientos tiene como consecuencia una infrautilización del canal, incluso sin la presencia de errores en el canal, con un rendimiento $\approx 20\%$ inferior al observado para la mejor configuración (semilla aleatoria, $K = 256$ y $q = 8$). Sin embargo, la utilización de esta alternativa (semilla aleatoria) limita el uso de la solución en un esquema *extremo a extremo*, impidiendo una posible futura recombinación de los mensajes por parte de los nodos intermedios, lo que sí sería posible con la transmisión explícita de los coeficientes.

A pesar que el valor de *throughput* ofrece una indicación absoluta del rendimiento, hay que tener en cuenta que la solución propuesta requiere que el receptor retenga los paquetes hasta que consiga decodificar el bloque completo, recuperando únicamente en ese instante el contenido de los K paquetes originales. De esta forma, la capa de aplicación del destino no recibe la información de manera fluida, sino que lo hace a ráfagas, con llegadas de K paquetes, lo que sucede tras cada evento de decodificación con éxito. Por este motivo, hay que tener en cuenta otros estadísticos, como el retardo medio entre decodificaciones consecutivas o la cantidad de información que se envía al nivel de aplicación por cada bloque recuperado. La Fig. 5 permite analizar ambos parámetros, en función del tamaño del bloque y del orden del cuerpo. En primer lugar, la Fig. 5a pone de manifiesto que pueden necesitarse hasta casi 500 milisegundos para decodificar un bloque de 256 paquetes, recibiendo la capa de aplicación más de 300 KB de información (Fig. 5b). Se observa además que el aumento en el valor de q repercute negativamente en la cantidad de datos que se recuperan por recepción exitosa, con una diferencia, para $K = 256$, de aproximadamente un 20% entre las configuraciones con $q = 1$ y $q = 8$. Esto podría tener una especial repercusión en ciertos tipos de servicio (videoconferencias, las llamadas de voz sobre IP), con estrictos requerimientos de tiempo real, *latencia* o *jitter*, lo que podría limitar la configuración a emplear por el esquema de codificación.

En un último análisis se ha configurado el enlace inalám-



(a) Transmisión explícita del vector de coeficientes



(b) Transmisión de la semilla aleatoria

Fig 4: *Throughput* en un enlace ideal

brico para que se produzcan pérdidas. Para ello, se ha utilizado un modelo de canal "sin memoria" en el que se va modificando la probabilidad de error por trama (*Frame Error Rate*, FER). Se asume que en los reconocimientos nunca se producen errores (lo que se corresponde con una situación real, ya que los mensajes más cortos son menos propensos a perderse). En este caso, se fijan en el transmisor unos valores de $q = 1$ y $K = 64$, que ofrecen un rendimiento cercano al óptimo, y presentan a la vez una eficiencia computacional mucho mayor⁴. También se estudiará la influencia del mecanismo de retransmisiones a nivel de enlace que incorpora el estándar *IEEE 802.11*, variando el número máximo de intentos por datagrama. La Figura 6 representa el comportamiento de la transmisión a medida que la calidad del enlace entre los nodos origen y destino se va deteriorando. Queda patente, en primer lugar, el beneficio de las retransmisiones de nivel de enlace al utilizar TCP, ya que permiten mitigar la pérdida de segmentos durante la transmisión; así, si se deshabilitan, el rendimiento cae de manera brusca en el momento en el que la presencia de errores comienza a ser apreciable. Al habilitar las retransmisiones, se puede ver que el sistema ofrecerá una resistencia mayor a la pérdida puntual de tramas, manteniendo un descenso prácticamente lineal, hasta que se llega a una FER

⁴A medida que aumenta el tamaño de la matriz y el orden de los cuerpos, más costosas se vuelven las operaciones necesarias en la decodificación. La versión actual del protocolo se apoya en una librería externa para el manejo del álgebra de cuerpos finito, *M4ri/M4rie* [17], que resulta ser más eficiente que otras alternativas que también se han analizado, como *FFLAS/FFPACK* [18] o *IT++* [19], entre otras.

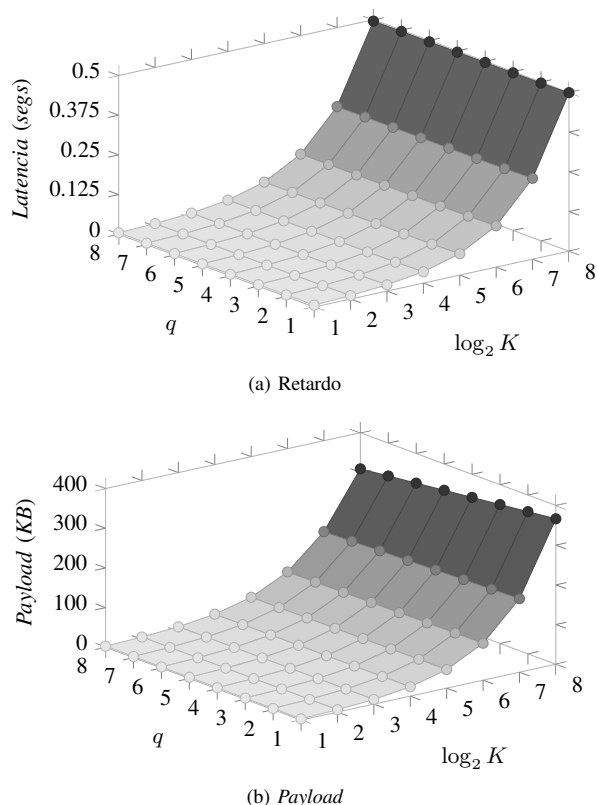


Fig 5: Dualidad entre retardo y cantidad de información que le llega a la aplicación tras la decodificación de un bloque

que excede la capacidad de recuperación de los niveles de red y de transporte, reduciendo de manera notable el rendimiento observado. A modo de ejemplo, en [8] se afirma que una pérdida de segmentos superior al 2,5% es suficiente para desestabilizar una conexión TCP. Por su parte, el esquema NC *intra-flujo* muestra un comportamiento muy diferente. En primer lugar, cualquiera de las configuraciones consigue ofrecer unos valores de *throughput* sensiblemente superiores a los vistos en el caso de TCP, con una ganancia que se acentúa a medida que las condiciones del canal empeoran. Sin embargo, también se pone de manifiesto que, en este caso, las retransmisiones de nivel de enlace no hacen sino perjudicar el rendimiento, que es superior si se deshabilitan. La explicación se encuentra en el propio algoritmo de *back-off* utilizado tras la pérdida de una trama, que duplica paulatinamente la ventana de contención utilizada a por el mecanismo de acceso al medio inalámbrico; sin embargo, como se ha mencionado anteriormente, la solución propuesta elimina la necesidad de recuperar un paquete codificado en concreto (carácter individual), siendo suficiente recibir K mensajes linealmente independientes “*cualesquiera*” para poder decodificar un bloque. Así, la espera adicional asociada al mecanismo de acceso al medio tendrá como consecuencia una infrutilización de los recursos disponibles, sin ningún beneficio adicional.

VI. CONCLUSIONES Y LÍNEAS FUTURAS

Una de las mayores limitaciones del protocolo TCP es su mala adaptación a las redes inalámbricas, sobre las que ofrece un rendimiento pobre, especialmente cuando la presencia de

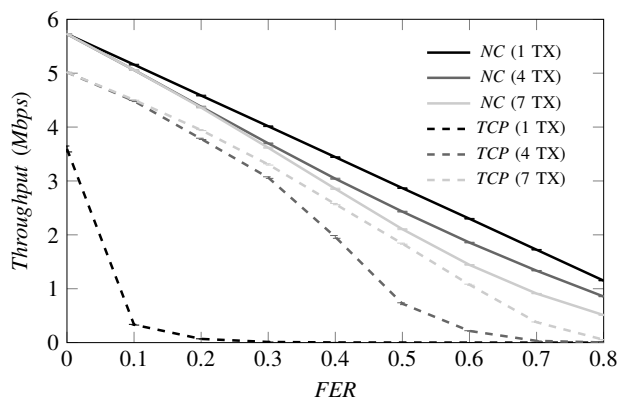


Fig 6: *Throughput* en un canal con errores

errores en el canal se vuelve apreciable. La causa principal de este problema es la agresividad de sus mecanismos de control de congestión ante la pérdida de segmentos, que siempre interpreta como una indicación de congestión en los nodos intermedios de la red.

Este trabajo presenta una aproximación novedosa, basada en la combinación de un esquema de codificación aleatoria (RLC) con el protocolo UDP. La unión de ambos esquemas garantiza la recepción ordenada de la información en el destino, ofreciendo por tanto un servicio similar al de TCP. Del mismo modo, la operación conjunta de ambos esquemas permite enmascarar la presencia de errores aleatorios durante una transmisión, asegurando una comunicación más robusta y estable. Para ello, una entidad NC, situada como una capa intermedia entre UDP e IP, almacenará en el nodo fuente los paquetes que envía la aplicación transmisora. Cuando disponga la información necesaria comenzará a enviar combinaciones lineales aleatorias de los paquetes que conforman un bloque, con coeficientes elegidos aleatoriamente sobre un cuerpo finito, hasta que la entidad par del equipo destino haya sido capaz de recuperar la información transmitida. Además, se han incluido una serie de soluciones *cross-layer* que permiten mejorar el comportamiento de la solución propuesta.

Se ha llevado a cabo una extensa campaña de simulación sobre la plataforma ns-3, para evaluar el comportamiento del protocolo propuesto sobre un enlace inalámbrico, comparando su rendimiento con el observado con una transmisión TCP clásica. En primer lugar, se ha observado la influencia de los parámetros de configuración que establecen la operación de la codificación: el tamaño del bloque y el orden del cuerpo finito sobre el que efectuar las operaciones algebraicas necesarias. Se ha observado que el número de transmisiones innecesarias disminuye a medida que se incrementan dichos parámetros, hasta hacerse prácticamente nulo. Sin embargo, la sobrecarga que supone transmitir los coeficientes utilizados para construir el paquete codificado puede penalizar el rendimiento, al disminuir el espacio disponible para transportar información útil, lo que puede solucionarse con la transmisión de la semilla utilizada en la generación de dichos coeficientes. En términos de *throughput*, se observa una mejora cercana al 20% sobre el de TCP en el mejor de los casos. También se ha comprobado que el aumento del tamaño de los bloques influye de manera directa en el retardo entre decodificaciones

consecutivas. Finalmente, se ha analizado la capacidad del protocolo propuesto para enmascarar los errores producidos en el canal radio, ofreciendo un rendimiento muy superior al de TCP, sobre todo a medida que las condiciones del canal se vuelven más hostiles; finalmente, se ha comprobado que la propiedad “rateless” del esquema de codificación elimina los posibles beneficios de los mecanismos de retransmisión del nivel de enlace, ya que su única consecuencia es el incremento de los tiempos de inactividad en la fuente durante el acceso al medio inalámbrico.

Tras el desarrollo de este trabajo, aún quedan abiertas numerosas líneas que serán tratadas en el futuro. En primer lugar, uno de los aspectos con mayor potencial se centra en la búsqueda de esquemas de codificación más eficientes computacionalmente, con soluciones que reduzcan el coste de los procesos de codificación/recodificación/decodificación. Otro de los elementos clave, ya adelantado en el documento, es la habilidad de los nodos intermedios de recombinar el contenido de los paquetes codificados que los atraviesan. Utilizando esta variación, sería interesante extender el análisis a redes malladas multi-salto, comprobando el beneficio que aportaría la recodificación adicional de los nodos intermedios. Además, aprovechando la cada vez mayor presencia de múltiples interfaces en los terminales, podría resultar beneficioso modificar el protocolo para que divida la transmisión por múltiples subflujos. Gracias a la propiedad “rateless” del esquema propuesto, se eliminaría la necesidad de incorporar complejos mecanismos de control de flujo o de reordenación en el destino, permitiendo incrementar notablemente el rendimiento neto ofrecido en el sistema. También se contempla el análisis de este protocolo sobre redes con alto *Bandwidth Delay Product* (BDP), en las que es necesario adaptar los mecanismos de control de congestión tradicionales de TCP. Esta problemática está tratada en el trabajo de Afanasyev et al. [20], en el que se describen las diferentes propuestas que se han realizado.

AGRADECIMIENTOS

Los autores querían expresar su agradecimiento al Gobierno de España por su financiación en el proyecto “*Connectivity as a Service: Access for the Internet of the Future*”, COSAIF (TEC2012-38574-C02-02).

REFERENCIAS

- [1] G. Xylomenos, G. Polyzos, P. Mahonen, and M. Saaranen, “Tcp performance issues over wireless links,” *Communications Magazine, IEEE*, vol. 39, no. 4, pp. 52–58, Apr 2001.
- [2] D. Gomez, S. Hassayoun, A. Herren, R. Aguero, and D. Ros, “Impact of network coding on tcp performance in wireless mesh networks,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, Sept 2012, pp. 777–782.
- [3] D. Gómez, E. Rodríguez, R. Agüero, and L. Muñoz, “Reliable communications over wireless mesh networks with inter and intra-flow network coding,” in *Proceedings of the 2014 Workshop on Ns-3*, ser. WNS3 '14. New York, NY, USA: ACM, 2014, pp. 4:1–4:8. [Online]. Available: <http://doi.acm.org/10.1145/2630777.2630781>
- [4] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [5] R. Koetter and M. Médard, “An algebraic approach to network coding,” in *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, 2001, p. 104.
- [6] S.-Y. Li, R. Yeung, and N. Cai, “Linear network coding,” *Information Theory, IEEE Transactions on*, vol. 49, no. 2, pp. 371–381, feb 2003.
- [7] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, “On randomized network coding,” in *In Proceedings of 41st Annual Allerton Conference on Communication, Control, and Computing*, 2003.
- [8] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, “XORs in the Air: Practical Wireless Network Coding,” *Networking, IEEE/ACM Transactions on*, vol. 16, no. 3, pp. 497–510, june 2008.
- [9] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, “Trading structure for randomness in wireless opportunistic routing,” in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '07. ACM, 2007, pp. 169–180. [Online]. Available: <http://doi.acm.org/10.1145/1282380.1282400>
- [10] M. Luby, “LT codes,” in *Proceedings of the 43rd Symposium on Foundations of Computer Science*, ser. FOCS '02. Washington, DC, USA: IEEE Computer Society, 2002, pp. 271–. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645413.652135>
- [11] A. Shokrollahi, “Raptor codes,” *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2551–2567, Jun. 2006. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2006.874390>
- [12] J. Sundararajan, D. Shah, M. Médard, S. Jakubczak, M. Mitzenmacher, and J. Barros, “Network Coding Meets TCP: Theory and Implementation,” *Proceedings of the IEEE*, vol. 99, no. 3, pp. 490–512, march 2011.
- [13] M. Kim, J. Cloud, A. ParandehGheibi, L. Urbina, K. Fouli, D. J. Leith, and M. Médard, “Network Coded TCP (CTCP),” *CoRR*, vol. abs/1212.2291, 2012.
- [14] J. Heide, M. Pedersen, F. Fitzek, and M. Medard, “A perpetual code for network coding,” in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, May 2014, pp. 1–6.
- [15] D. E. Lucani, M. V. Pedersen, J. Heide, and F. H. P. Fitzek, “Fulcrum network codes: A code for fluid allocation of complexity,” *CoRR*, vol. abs/1404.6620, 2014. [Online]. Available: <http://arxiv.org/abs/1404.6620>
- [16] “The ns-3 network simulator,” 2008, <http://www.nsnam.org/>.
- [17] “M4RI(e)- Linear Algebra over GF(2) and GF(2+),” <http://m4ri.sagemath.org/>.
- [18] “FFLAS-FFPACK. Finite Field Linear Algebra subroutines package,” <http://www-ljk.imag.fr/membres/Jean-Guillaume.Dumas/FFLAS/index.html>.
- [19] “IT++ Mathematical library,” <http://itpp.sourceforge.net/>.
- [20] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, “Host-to-host congestion control for tcp,” *Communications Surveys Tutorials, IEEE*, vol. 12, no. 3, pp. 304–342, Third 2010.

A Dynamic Multimetric Weights Distribution in a Multipath Routing Protocol using Video-Streaming Services over MANETs

Ahmad Mohamad Mezher¹, Cristhian Iza Paredes¹, Carolina Tripp-Barba², Mónica Aguilar Igartua¹

¹ Department of Network Engineering.
Universitat Politècnica de Catalunya (UPC)
C/ Jordi Girona 1-3, 08034 Barcelona, Spain.

² Faculty of Informatics.
Universidad Autónoma de Sinaloa (UAS)
Mazátlan, 82107, Mexico.

[ahmad.mezher, cristhian.iza, maguilar]@entel.upc.edu, ctrippp@uas.edu.mx

Abstract—Mobile ad hoc networks (MANETs) are infrastructureless networks formed by wireless mobile devices. Routing in MANETs is a process where the best path is selected to forward packets. However, when a MANET is heavily loaded, single path routing may have congestion problems. To cope with this issue, multipath routing protocols were proposed to improve single path routing in terms of packet delivery ratio and quality of service (QoS). Normally, QoS metrics are used to classify paths to choose paths which satisfy the QoS user requirements. Besides, paths usually are classified with equal metric weights which may not be an optimal way. In this paper, we intend to design a new way to compute the weight values for metrics so that paths are classified in a better way. Each time we want to classify paths, our proposal seeks to give a proper value to each metric weight that depends on the metric current value with respect to an average reference previously updated. We consider that depending on the variation of the metric value, a corresponding weight to this current variation must be given. To achieve that, a smart algorithm has been designed. Finally, paths are classified taking into account our algorithm to calculate all weights' values. Simulations have been done to show the benefits of our proposal where interfering traffic and mobility of the nodes are present.

Keywords—Mobile ad hoc networks, adaptive multipath routing, video-streaming services, QoS metric weights.

I. INTRODUCTION

A Mobile Ad hoc NETWORK (MANET) is a group of self-organized wireless mobile nodes (MNs) able to communicate with each other without the need of any fixed network infrastructure nor centralized administrative support. MANETs suffer from link breakages due to nodes that move and have limited battery life, which produce frequent changes in the network topology. In addition, the transmission range in such mobile devices is limited, so multi-hop paths routing protocols will be needed. Each MANET node can operate as a terminal host and as a forwarding node. MANETs should adapt dynamically to be able to maintain communications despite of all these issues [1].

MANETs have attracted much attention from the research community over the last years and important technical advances have risen. These multi-hop networks are foreseen as a new kind of next generation access networks, where multimedia services will be demanded by end users from their wireless devices everywhere. In different situations, users

can spontaneously form an infrastructureless ad hoc network to share their resources and contents. Besides, multimedia services require Quality of Service (QoS) provision. The QoS provision over these networks is a challenge due to the special characteristics of MANETs, such as mobility, dynamic network topology, energy constraints, infrastructureless and variable link capacity. Thus, instead of using fixed network configuration parameters, a better solution would be to adjust the framework parameters according to the current environmental conditions.

Our research focuses on finding an algorithm capable to give each QoS metric a weight to compute a multimetric value by our multipath routing protocol that would help to classify paths in a better way. This algorithm should further enhance the performance of the service.

The rest of the paper is structured as follows. Section II introduces some relevant related work. Section III presents the basics of our framework. In section IV we explain the features of our multipath routing protocol. Section V gives a brief explanation of our proposal. Simulation results are shown and analyzed in section VI. Finally, conclusion and future work are given in section VII.

II. RELATED WORK

Routing is the process of selecting the best path or paths in a network through which data will be forwarded. Forwarding paths are selected usually seeking to improve the service performance and are not necessarily the shortest ones. Several works such as [2], [3], [4], [5], [6] presented interesting approaches to transmit video flows over multipath frameworks. The authors in [2] analyzed the topic of multipath routing for multiple description video coding in wireless ad hoc networks. They found that genetic algorithms (GAs) are effective to address this type of cross-layer optimization problems. They demonstrate the superior performance of their GA-based approach using numerical results. In [3] the authors proposed an optimal routing algorithm to distribute video over multiple paths seeking to minimize congestion and to improve the video quality. The authors in [4] designed an adaptive multipath multimedia dynamic source routing

(MMDSR) protocol able to self-configure dynamically depending on the state of the network while taking into account the special features of the video frames to distribute them over the multipath scheme. The authors in [5] proposed an extension of multipath optimized link state routing protocol (MP-OLSR) named fuzzy-based quality of service MP-OLSR (FQ-MP-OLSR) integrating two fuzzy systems. The first is used to calculate a multi-constrained QoS metric based on delay, throughput and signal to interference plus noise ratio (SINR) while the second is applied to adapt cost functions used to penalize paths previously computed by the Dijkstra's algorithm. Simulation results showed that FQ-MP-OLSR achieves a significant improvement in terms of QoS and quality of experience (QoE). The work in [6] presented a QoS-aware routing framework combining three QoS mechanisms, i.e., cross-layer communication mechanism, session admission control and QoS-aware multipath routing to achieve an efficient video transmission over MANETs. Simulation results showed the improvement obtained in terms of video quality. Traditionally, all the works that use several metrics weight them equally. However, we think that an algorithm to better calculate those weights throughout time is still needed. In this present work, for our purpose of updating the QoS weight metrics values in MANETs, we will further improve our previous proposal MMDSR [4]. MMDSR is a QoS-aware self-configured multipath routing protocol that dynamically adapts to the changing environment. In this work, the weights of the QoS metrics will be self-configured. We will design a new way to compute and update those weights so that paths would be better arranged according to their current QoS metrics and the current state of the environment. In this way, each time the forwarding algorithm needs to classify paths, a proper weight value to each metric will be updated depending on the current state of the network. In the next section, we summarize the basics of our framework.

III. BASICS OF THE FRAMEWORK

The multipath routing scheme of MMDSR is based on the DSR (Dynamic Source Routing) protocol [7] to find available paths from source to destination. Video is distributed using RTP/RTCP (Real-time Transport Protocol/Real-time Control Protocol) [8] over UDP as transport protocols. Our system uses a layered MPEG-2 VBR coding of the video flow, which is formed by sets of 15 frames, called GoP (Groups of Pictures), see Fig. 1. A GoP has three types of frames: I, P and B, and has a unique frame-pattern in a video repeated in each GoP. I (Intra) frames encode spatial redundancy, they form the base layer, provide a basic video quality and carry the most important information for the decoding process at the receiving side. At decoding time, if the corresponding I frame were not available, the whole GoP would be lost. P (Predicted) and B (Bi-directional) frames carry differential information from preceding (P) or preceding and posterior (B) frames, respectively. Considering these characteristics, we assign different priorities to the video frames according to their importance within the video flow. Accordingly, I frames should have the highest priority, P frames the medium priority and B frames the lowest one.

In the MAC (Media Access Control) layer, we use the IEEE 802.11e [9] standard, which provides QoS support. It consists

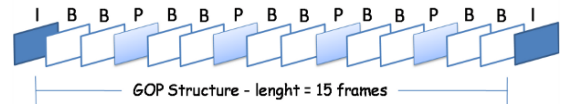


Fig. 1. MPEG-2 GoP structure.

of four different Access Categories (AC), each with their own configuration parameters, i.e. contention window minimum (CW_{min}) and maximum (CW_{max}), arbitration interframe space (AIFS) and transmission opportunity (TxOp). Each packet arrives at the MAC layer from the higher layers with a specific priority value and is mapped into the proper AC. We defined the mapping of the different packets into each one of the four AC_s as follows:

- AC0: signaling.
- AC1: high priority packets (I frames).
- AC2: medium priority packets (P frames).
- AC3: low priority packets (B frames + other best effort traffic).

IV. MULTIPATH MULTIMEDIA DYNAMIC SOURCE ROUTING (MMDSR)

In this section, we will give a brief summary of the main features of the framework, whose complete description was presented in [4]. In this present paper, we aimed to design a smart algorithm to update dynamically the weight values of several QoS metrics to compute a multimetric QoS value. This way, paths could be classified in a better way to set the multipath scheme.

A. Multipath routing scheme

MMDSR is a multipath routing protocol that uses the standard DSR as the engine to search available paths. MMDSR uses up to three paths through which the three types of video frames will be sent. As Fig. 2 shows, traditionally the most important video frames (I frames) would be sent through the best available path (high priority), P frames through the second best path (medium priority) and B frames through the third best path (low priority). In [10], [11], authors obtained that there are notable benefits in arranging two or three paths to transmit the video frames. However, arranging more than three paths in a multipath scheme does not provide a big improvement, while increases unnecessarily the management.

The user requirements are considered using QoS parameters knowing their threshold values to provide the negotiated image quality. We use the following parameters: minimum expected bandwidth (BW_{min}), the maximum percentage of data losses (L_{max}), the maximum delay (D_{max}) and the maximum delay jitter (J_{max}).

$$customer_req \equiv \{BW_{min}, L_{max}, D_{max}, J_{max}\} \quad (1)$$

B. MMDSR control packets

All decisions such as the tuning of configuration parameters or the path selection are operated from the source.

MMDSR protocol periodically discovers A available paths between source and destination by sending monitoring *Probe Message* (PM) packets. After that, a *Probe Message Reply*

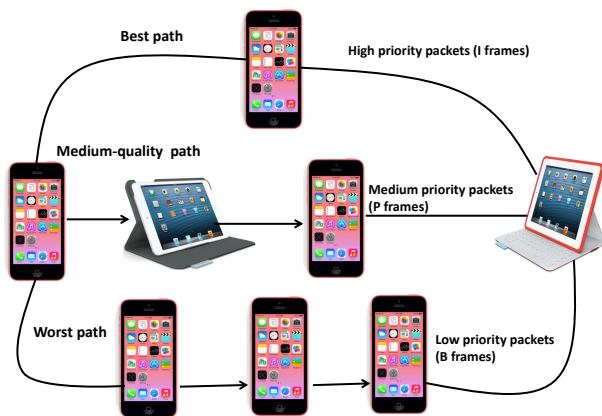


Fig. 2. Multipath routing scheme using three paths.

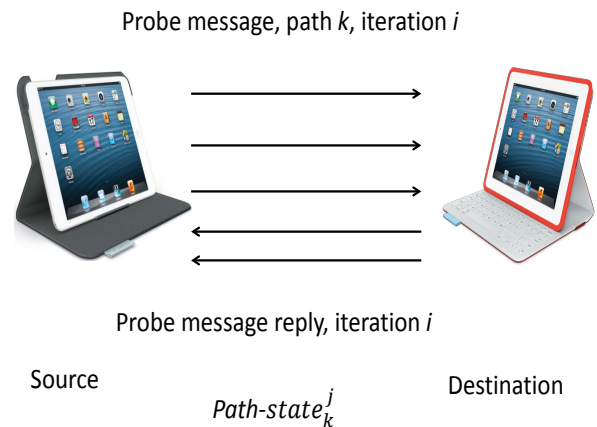


Fig. 3. PM and PMR packets.

(PMR) packet is generated at destination to carry the collected information about the quality of the available paths. It is important to mention that the reduced size of these packets (around 64 bytes, depending on the number of paths found out) and the low frequency of sending them (every 3-18 s depending on the network state, see eq. (4)) makes the incurred overhead almost negligible. Fig. 3 shows the PM and PMR packets that are periodically interchanged between source and destination.

Finally, a score is given to each one of the paths after analyzing the feedback information at the source node to classify them accordingly. The quality parameters of the paths will be compared to certain thresholds and then the source selects three paths (if they exist) to compose the multipath scheme. $path-state$ is a vector with all quality parameters calculated for each one of the available paths:

$$path - state_k^i \equiv \{D, J, H, RM, MM\}_k^i \quad (2)$$

where i is the iteration number of the algorithm and k refers to each one of the paths (with $k \leq A$). The other parameters are: delay (D_k^i), delay jitter (J_k^i), hop distance (H_k^i), reliability Metric (RM_k^i) calculated from the SINR (Signal to Interference plus Noise Ratio) of the links involved in each path, and Mobility Metric MM_k^i calculated from the relative mobility of the neighboring nodes within each path.

To refresh the paths, this process is repeated periodically (i.e., each $T_{routing}$ s) due to the topology of MANETs that varies and that produces link breakages. This routing period depends on the network state, as it is shown in section IV-C.

C. MANET self-configuration

Here, we merely point out the basics of the self-configuration operation. For further details please see [4]. Our framework monitors the current state of the network and in case of changes, the algorithm modifies some configuration parameters such as the routing period of the algorithm and the thresholds to classify paths. $NState$ has information about the global network state and is updated by the algorithm iteration by iteration. $NState$ is computed as follows:

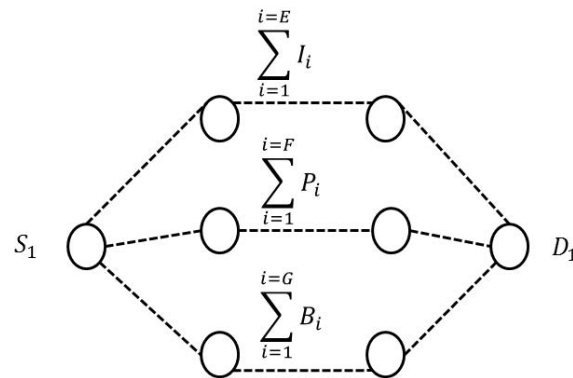


Fig. 4. Fixed strategy to allocate resources. E , F and G represent the number of I, P and B frames sent through the best, medium and worst path, respectively.

$$NState^i = w_{RM} \cdot \overline{RM^i} + w_{MM} \cdot \overline{MM^i} + w_D \cdot \overline{D^i} + w_J \cdot \overline{J^i} + w_H \cdot \overline{H^i} \quad (3)$$

In eq. (3) upper bars denote averages and the w s are appropriate weights that sum one. These weights values could be equal (i.e., $w_{RM} = w_{MM} = w_D = w_J = w_H = 1/5$) or different from each other (using our proposed algorithm). When the source receives the feedback from the network by means of a PMR packet, it calculates the $NState$ using eq. (3).

As $NState$, the routing period ($T_{routing}$) to refresh the multipath scheme also varies dynamically and it is calculated using eq. (4).

$$T_{routing}^{j+1} = \gamma \cdot NState^j + \theta \quad (4)$$

To obtain the previous equation, a high number of simulations were conducted under a wide range of situations where

the network performance was good, normal and bad. For the scenario under consideration, the obtained values were $\gamma = 10$ and $\theta = 3$.

Till now, we have presented the basics of a QoS-aware adaptive multipath routing protocol. Next, we introduce the novel algorithm to update the metrics weights to further improve the performance of video streaming services over MANETs.

V. ALGORITHM TO UPDATE THE WEIGHTS OF THE METRICS

A. Motivation

Several multipath routing protocols for MANETs based on end-to-end path have been proposed in the literature. Some of them use several metrics to arrange the paths and select the best ones [4], [11]. All those proposals give equal weights (i.e., they have the same importance) to the QoS metrics. Nonetheless, a smart algorithm able to give each metric a proper weight depending on a defined strategy is still needed. For this reason, we propose an algorithm that seeks to improve the classification of the discovered paths. As a consequence, paths could be classified in a better way. We think that if in each time we want to classify our paths into best, medium and worst, weights of the metrics implemented in the multipath routing protocol could take different values than $\frac{1}{N}$ (being N the number of metrics). We propose an algorithm to compute the weights of the metrics as a function of the variation of the values of the normalized metrics with respect to the average of those values. These average values are obtained from conducting a high number of simulations. For instance, at each $T_{routing}$ seconds, if we see that a metric value differs a lot from its average value, then we will give this metric more importance (i.e., a higher weight), since this metric will allow us to better differentiate the quality of the paths.

B. Algorithm Description

Suppose that we have N metrics, M_1, M_2, \dots, M_N with their corresponding weights w_1, w_2, \dots, w_N where $\sum_{i=1}^N w_i = 1$. Previously, we carried out 100 simulations and we got the average value for each metric M_j , named A_j . V_j is the current value for metric M_j in each $T_{routing}$ seconds. Each $T_{routing}$ seconds, sources reclassify paths as best, medium and worst by using new weights that are calculated according to our algorithm.

Let us denote R_k as the variation made in each $T_{routing}$ seconds (each i iteration) for each metric in path k , such as:

$$R_k = \begin{cases} R_{1,k} = \left| \frac{V_{1,k} - A_{1,k}}{V_{1,k}} \right| \\ R_{2,k} = \left| \frac{V_{2,k} - A_{2,k}}{V_{2,k}} \right| \\ \vdots \\ R_{N,k} = \left| \frac{V_{N,k} - A_{N,k}}{V_{N,k}} \right| \end{cases} \quad (5)$$

After that, we will have a vector $R_k = [R_{1,k}, R_{2,k}, \dots, R_{N,k}]$. Suppose that the maximum value found in vector R_k is $R_{max,k} = R_{x,k}$ where $x \in [1, N]$. Now, we bound vector R_k to be between 0 and 1 (i.e., $0 \leq R \leq 1$) and this new vector will be named S_k .

$$S_k = \begin{cases} S_{1,k} = \frac{R_{1,k}}{R_{x,k}} \\ S_{2,k} = \frac{R_{2,k}}{R_{x,k}} \\ \vdots \\ S_{x,k} = \frac{R_{x,k}}{R_{x,k}} = 1 \\ \vdots \\ S_{N,k} = \frac{R_{N,k}}{R_{x,k}} \end{cases} \quad (6)$$

Here, we calculate a parameter α_k value using eq. (7):

$$\alpha_k = \frac{1}{\sum_{i=1}^N S_{i,k}} \quad (7)$$

Therefore, the new normalized vector weights W_k for the metrics will be:

$$W_k = \begin{cases} W_{1,k} = S_{1,k} \cdot \alpha_k \\ W_{2,k} = S_{2,k} \cdot \alpha_k \\ \vdots \\ W_{x,k} = S_{x,k} \cdot \alpha_k = \alpha_k \\ \vdots \\ W_{N,k} = S_{N,k} \cdot \alpha_k \end{cases} \quad (8)$$

In the next section, we will define the metrics used in the MMDSR protocol.

C. Metrics used in the MMDSR routing protocol

In this section, we will define the metrics used to score the paths in our multipath routing protocol. The considered metrics are: delay, delay jitter, number of hops, reliability metric and mobility metric. Regarding the score of each metric, we selected a negative exponential function to penalize those paths with bad values in that metric. This way, we highlight the best candidate paths, so that they have more chances to be selected as one of the three paths needed to form the multipath scheme. As a consequence, the best path will get a score notably higher than the others.

1) *Delay and delay jitter*: The shorter the value of the delay, d_k , the better the path will be classified. Also, for the delay jitter the shorter the value of the delay jitter, dj_k , the better the path will be scored.

Once the PMR packet arrives at the source node, this node gets the values for delay and jitter which have been sampled in each path. Having these sampled values, the source node computes the delay d_k^i , delay jitter dj_k^i for each k path in the current i iteration. We apply EWMA (Exponentially Weighted Moving Average) filters, with a low value for the coefficient (i.e., $\beta = 0.125$) to give more relevance to the historical values in front of the instantaneous ones. As a consequence, parameters will evolve in a smooth way:

$$d_k^i = (1 - \beta) \cdot d_k^{i-1} + \beta \cdot d_k^{i-sample} \quad (9)$$

$$dj_k^i = (1 - \beta) \cdot dj_k^{i-1} + \beta \cdot dj_k^{i-sample} \quad (10)$$

Then, we compute $u_{1,i,k}$ and $u_{2,i,k}$ as the score for path k in each iteration i for delay and delay jitter, respectively. As we have mentioned before, we apply a negative exponential function as shown in eq. (11) and eq. (12).

$$u_{1,i,k} = e^{-d_k^i} \quad (11)$$

$$u_{2,i,k} = e^{-dj_k^i} \quad (12)$$

2) *Number of hops, h_k^i* : We calculate the number of hops of each path k (i.e. Hop Metric h_k^i). This way, we assign higher scores to shorter paths. Shorter paths are preferred, as fewer losses will take place due to contention for the medium.

In this case, the shorter the value of the number of hops, h_k^i , the better the path will be classified, because we prefer a path that has lower hops. The score related to the number of hops for path k in iteration i is computed as shown in eq. (13).

$$u_{3,i,k} = e^{-h_k^i} \quad (13)$$

3) *Reliability metric*: We propose to compute a performance measure of the whole quality of a path by taking SINR (Signal-to-Interference plus Noise Ratio) measures between consecutive nodes within path k . For each one of the i iterations of the algorithm and each k path, we obtain the $SINR_j^{k,i}$ values of each j node with respect to the $j - 1$ node on the downstream path from source to destination.

We compute the mean qualification of the whole path k , RM_{mean}^k , applying the geometric average of all the partial $SINR_j^{k,i}$ qualifications of each node j of the path k . L_k^i corresponds to the number of hops in each path k . This way, RM_{mean}^k is computed as follows:

$$RM_{mean}^k = \sqrt[L_k^i]{\prod_{j=1}^{j=L_k^i-1} SINR_j^{k,i}} \quad (14)$$

In this case, the larger the value of the reliability metric value, RM_{mean}^k , the better the path will be classified.

$$u_{4,i,k} = e^{-\left(\frac{1}{RM_{mean}^k}\right)} \quad (15)$$

4) *Mobility metric*: Each node X detects the received signal power $RxPr_{Y \rightarrow X}$ from its neighbour Y (which belongs to one of the available paths from source to destination) from successive packet transmissions (periodic Hello messages). Once this value returns to the original node of the Hello Message, each local qualification for the relative mobility metric is computed as follows:

$$M_X^i = E \left[\log_{10} \left(\frac{RxPr_{Y \rightarrow X}^i}{RxPr_{Y \rightarrow X}^{i-1}} \right)^2 \right] \quad (16)$$

A high value for M_X^i shows that X moves more than their neighbours, while a low value means that the X node is relatively less mobile than its neighbours. A path whose nodes have a lower relative mobility will be preferred in front of other paths whose nodes have a high global value of mobility. The next time a Probe Message arrives at node X, the node will append the current value of its mobility metric qualification, M_X^i . After this, the destination computes the

global qualification of the mobility metric for that path k in this way: for each path k , we obtain the M_X^i values of each node X within that path using eq. (16). Finally, we compute the arithmetic mean from the M_X^i values to obtain a measure of the mobility of path k , during iteration i of the algorithm, MM_k^i , as shown in eq. (17).

$$MM_k^i = \left(\frac{\sum_{j=1}^{j=L_k^i-1} M_X^i}{L_k^i} \right) \quad (17)$$

In this case, the smaller the value of the mobility metric value, MM_k^i , the better the path will be classified.

$$u_{5,i,k} = e^{-MM_k^i} \quad (18)$$

D. Path classification decision

When the source finds out all the possible paths between source and destination and wants to classify them, it evaluates and assigns a total qualification to each path. We assign the weights ($w_{1,i,k}, w_{2,i,k}, w_{3,i,k}, w_{4,i,k}, w_{5,i,k}$) to each metric ($u_{1,i,k}, u_{2,i,k}, u_{3,i,k}, u_{4,i,k}, u_{5,i,k}$) in the qualification to each path k , iteration i . We propose a weighted geometric average metric as seen in eq. (19). A global geometric score is used, because it is less sensitive than the arithmetic metric in the extreme values of the metric components. Due to the fact that all the metrics are exponential functions with the same base, the best path will be the one that has the greatest sum of all the exponents.

$$\begin{aligned} \bar{u}_{i,k} &= \prod_{j=1}^5 u_j^{w_j} \\ \bar{u}_{i,k} &= u_1^{w_1} \cdot u_2^{w_2} \cdot u_3^{w_3} \cdot u_4^{w_4} \cdot u_5^{w_5} \end{aligned} \quad (19)$$

We give a score $\bar{u}_{i,k}$ to each path discovered by the routing protocol. Therefore, the three best paths are those that have the highest multi-metric value obtained applying our algorithm.

Algorithm 1 Ordering paths from the best till the worst one.

Require: Input data (opt.routes, Score, available paths = A)

```

1: for (i = 1; i < A; i++) do
2:   for (j = 1; j < A-1; j++) do
3:     if Score ([j]) < Score ([j-1]) then
4:       tmp.routes ([j]) = opt.routes ([j])
5:       opt.routes ([j]) = opt.routes ([j+1])
6:       opt.routes ([j+1]) = tmp.routes ([j])
7:       tmp = score ([j])
8:       score ([j]) = score ([j+1])
9:       score ([j+1]) = tmp
10:    end if
11:  end for
12: end for

```

Fig. 4 shows the proposed architecture. In this figure, we see one connection (S_1-D_1) and three paths. Anyway, it is possible to apply this proposed architecture to any MANET independently from the number of connections, nodes and paths.

Algorithm 1 shows how we arrange each path with its corresponding score from the highest to the lowest. In this

Table I
SIMULATION SETTINGS OF THE MANET SCENARIO.

Area	520x520 m ²
Number of nodes	50
Average node speed	2 m/s
Transmission range	120 m
Mobility Pattern	Random Waypoint
MAC specification	IEEE 802.11e, EDCAF
$CW_{i,min}, i \in (0,1,2,3)$	(7, 15, 31, 63)
$CW_{i,max}, i \in (0,1,2,3)$	(15, 31, 1023, 1023)
$AIFS_i, i \in (0,1,2,3)$	(34 μ s, 43 μ s, 52 μ s, 61 μ s)
$(BW_{min}, L_{max}, D_{max}, J_{max})$	(50 Kbps, 35 %, 0.125 s, 0.004 s)
Nominal bandwidth	11 Mbps
Simulation time	300 seconds
Video encoding	MPEG-2 VBR
Video bit rate	150 Kbps
Video sources	1
Video sequence sent	Blade Runner
Routing protocol	Multipath Routing protocol
Transport protocol	RTP/RTCP/UDP
Maximum packet size	1500 Bytes
Multipath scheme	$K=3$ paths
Weighting values (equation (3))	1/5 or variable weights using algorithm 1
Queue sizes	50 packets
Interfering CBR traffic	300 Kbps
Channel noise	-92 dBm
Mobility pattern generator	Bonnmotion

way, the first three best paths (if they exist) will be the best, the medium and the worst path to be used by the source.

By default, nodes always try to send the most important video frames through the best available path discovered by the multipath routing protocol. This means that I frames will be sent through the best path, whereas the least important frames (i.e., B frames) will be sent through the worst one. If only one path satisfied user's requirements, the source would send I, P and B frames through that single founded path.

VI. SIMULATION RESULTS

Our proposal was implemented in the open source network simulator ns-2 [12] where we conducted simulations to evaluate the benefits of our approach. The MANET scenario was generated with the Bonnmotion tool [13]. Interfering CBR traffic was generated to constrain the paths. The simulation settings of the scenario are shown in table I.

The scenario used to test the proposal consists of a set of 50 motion nodes distributed in a MANET of 520x520 m². The transmission range of the nodes is 120 m. Average node speed is 2 m/s. Video flows are transmitted from node S_1 to D_1 . The source sends frames I, P and B through the best, medium and worst path, respectively after classifying them according to the figure depicted in Fig. 2.

All the figures present confidence intervals (CI) of 90% obtained from ten simulation per point, each with a different Bonnmotion [13] scenario. In the following, results of losses, average jitter delay and average end-to-end delay are shown for the case of using equal weights compared to the case of using variable weights according to our algorithm.

Fig. 5 shows the average percentage of packet losses when using our algorithm compared to the case of non using it. We can see how including our algorithm, the average video packet losses are reduced from 12.1% to 8.8%. This decrement is due to the variable values distribution of metrics that give each

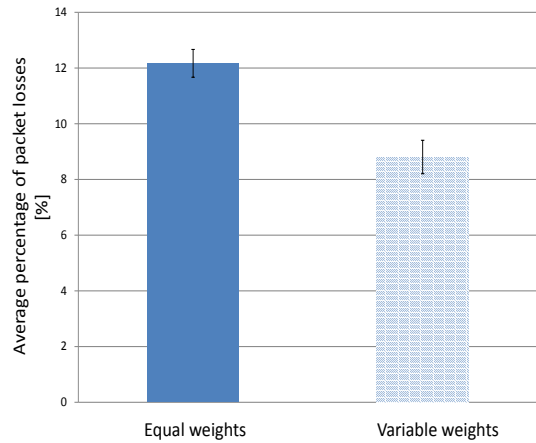


Fig. 5. Average percentage of packet losses.

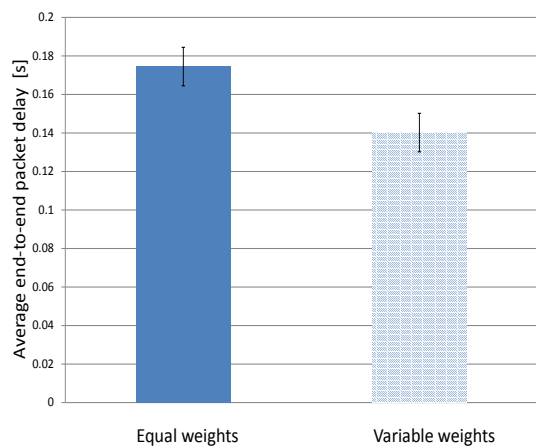


Fig. 6. Average end-to-end packet delay (s).

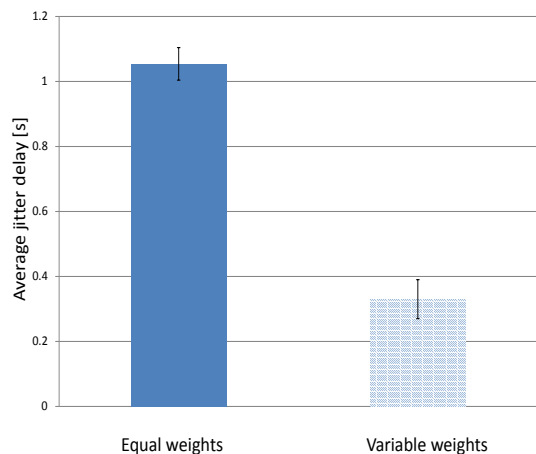


Fig. 7. Average jitter delay (s).

metric its importance in each iteration i , so that the paths are better chosen.

Fig. 6 depicts the average end-to-end delay. We can see that including our algorithm produces an improvement in terms of delay compared to the case of using equal weights. It is reduced from 0.17 s to 0.14 s.

Fig. 7 depicts the average jitter delay suffered by the packets. Also, including our algorithm produces an improvement in terms of jitter delay. It is reduced from 1.05 s to 0.33 s.

Fig. 8 shows the percentage of mismatched path classification between our scheme and the case of using equal weights for each one of the ten simulations. For example, using equal weights values, path k is chosen as best one while using our algorithm, this path k is not chosen the best one. This is considered as a mismatched case. We see that the percentage of mismatched varies between 14.26% and 66%. It is calculated using eq. (20).

$$\text{Mismatched percentage [\%]} = 100 \cdot \left(\frac{a}{b} \right) \quad (20)$$

$$\text{where } a = \left(\frac{\sum_{i=1}^{\lfloor \frac{\text{Simulation time}}{T_{\text{routing}}} \rfloor} G_i - \sum_{i=1}^{\lfloor \frac{\text{Simulation time}}{T_{\text{routing}}} \rfloor} C_i}{\sum_{i=1}^{\lfloor \frac{\text{Simulation time}}{T_{\text{routing}}} \rfloor} G_i} \right)$$

$$\text{and } b = \sum_{i=1}^{\lfloor \frac{\text{Simulation time}}{T_{\text{routing}}} \rfloor} 1.$$

In eq. (20), G_i is the total number of paths that satisfy the QoS requirements in iteration i ; C_i indicates how many paths classifications coincided with the usual path classification using equal weights and the classification using our algorithm. For example, in a specific iteration, using equal weights, path 3 is classified as the best, path 1 as the medium and path 2 as the worst, while using our algorithm path 1 is classified as the best where path 3 is the medium and path 2 is classified as the worst. In this case, we have 2 different classification out of 3. This means that in this iteration we have a difference of 66.6% in the path classification for this iteration. We average these percentages for all the iterations along the simulation.

Our multipath routing protocol MMDSR plus our proposal to distribute the values of the weights have shown how the global benefits of the network improve. Our algorithm produces lower video frame losses and thus a higher received video quality. In addition, we would like to highlight that in bad network conditions is where our algorithm would produce better benefits compared to the traditional case of using equal weights. This is because in excellent conditions of the network (i.e., T_{routing} high), paths will not be refreshed so often and as a consequence the same weights values will be used longer.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we designed a proposal to update the weight values for metrics in a multipath routing protocol to provide video-streaming services in mobile ad hoc networks. As a consequence, paths would be classified in a better way. Each time the routing protocol needs to classify paths, our proposal seeks to give a proper value to each metric depending on its

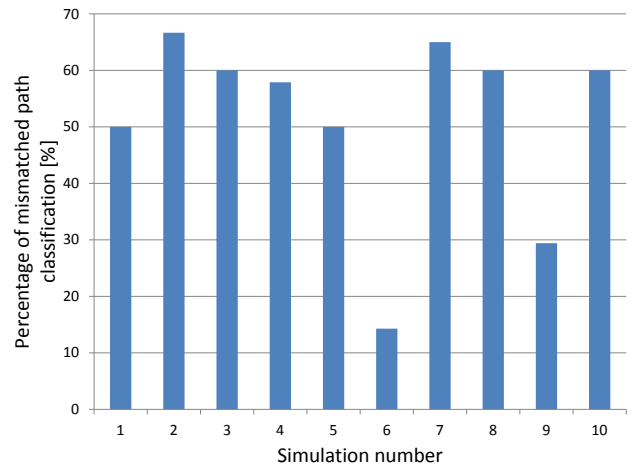


Fig. 8. Percentage of mismatched path classification between using our scheme with respect to the case of equal weights.

variation with respect to an average value previously updated. We consider that depending on the variation of the current value of a metric, the corresponding weight to this current variation could be computed.

Simulation results show the benefits of our proposal, outperforming the results compared to the case of non using it, in terms of average packet losses, average end-to-end packet delay and average jitter delay. Our proposal makes the network more efficient and achieves a higher degree of QoS provided to the users.

As a future work, we are planning to apply our proposal in vehicular ad hoc networks (VANETs), where the network topology is very dynamic due to the high speed of vehicles. In this case the design of a self-organized framework able to update its configuration parameters has special importance.

ACKNOWLEDGEMENTS

This work was partly supported by the Spanish Government through projects TEC2013-47665-C4-1-R (EMergency Response In Smart COMMunities, EMRISCO), TEC2014-54435-C4-1-R (INCident monitoRING In Smart COMMunities, QoS and Privacy, INRISCO), AGAUR Information Security Group (ISG) project - 2014 SGR 1504 and the research group Signals and Systems (UAS-CA-169) of the Autonomous University of Sinaloa (Mexico). Ahmad Mezher is the recipient of a FI-AGAUR grant, from the Government of Catalonia. Cristian Iza is recipient of a grant from Secretaria Nacional de Educación Superior, Ciencia y Tecnología SENESCYT.

REFERENCES

- [1] A. Boukerche, *Algorithms and Protocols for Wireless, Mobile Ad Hoc Networks*. Wiley-IEEE Press, 2008.
- [2] X. C. H. S. S. Mao, Y.T. Hou and S. Midkiff, "Multipath routing for multiple description video in wireless ad hoc networks," in *Proc. IEEE INFOCOM*, March 2005, pp. 740–750.
- [3] E. S. X. Zhu and G. Brend, "Congestion-distortion optimized video transmission over ad hoc networks," *Signal Processing: Image Communication*, vol. 20, pp. 773–783, 2005.
- [4] M. Aguilar and V. Carrascal, "Self-configured multipath routing using path lifetime for video-streaming services over ad hoc networks," *Computer Communications*, vol. 33, pp. 1879–1891, 2010.

- [5] A. Boushaba, A. Benabbou, R. Benabbou, A. Zahi, and M. Oumsis, "Intelligent multipath optimized link state routing protocol for qos and qoe enhancement of video transmission in manets," *Networked Systems, Lecture Notes in Computer Science*, vol. 8593, pp. 230–245, 2014.
- [6] C. Lal, V. Laxmi, and M. Gaur, "Bandwidth-aware routing and admission control for efficient video streaming over manets," *Wireless Networks*, vol. 21, pp. 95–114, 2015.
- [7] "RFC 4728, the dynamic source routing protocol (dsr) for mobile ad hoc networks for ipv4," 2007. [Online]. Available: <http://www.ietf.org/rfc/rfc4728.txt>
- [8] "RFC 3550, rtp: A transport protocol for real-time applications," 2003. [Online]. Available: <http://www.ietf.org/rfc/rfc4728.txt>
- [9] "IEEE 802.11e standard with quality of service enhancements," 2005. [Online]. Available: <http://standards.ieee.org/getieee802/download/802.11e-2005.pdf>
- [10] V. Loscri, F. D. Rango, and S. Marano, "Performance evaluation of ondemand multipath distance vector routing protocol over two mac layers in mobile ad hoc networks," in *Proc. 1st International Symposium on Wireless Communication Systems*, September 2004.
- [11] A. Mezher, M. A. Igartua, L. J. de la Cruz Llopis, E. P. Segarra, C. Tripp-Barba, L. Urquiza-Aguiar, J. Forné, and E. S. Gargallo, "A multi-user game-theoretical multipath routing protocol to send video-warming messages over mobile ad hoc networks," *Sensors Networks*, vol. 15, pp. 9039–9077, 2015.
- [12] "The network simulator, ns-2." [Online]. Available: <http://nsnam.isi.edu/nsnam/>
- [13] "Bonnmotion, a mobility scenario generation and analysis tool," 2005. [Online]. Available: <http://web.informatik.uni-bonn.de/IV/Mitarbeiter/dewaal/BonnMotion/>

Joint impact of CQI feedback delay and CQI reporting rate on channel-aware scheduling

Jose Oscar Fajardo, Ianire Taboada, Fidel Liberal

University of the Basque Country UPV/EHU, Alameda Urquijo s/n, 48013 Bilbao, Spain
{joseoscar.fajardo, ianire.taboada, fidel.liberal}@ehu.es

Abstract—Radio Resource Management (RRM) function is a critical element in multi-user mobile networks. Nowadays, RRM is implemented at the 4G radio edge in order to enhance channel-awareness and thus to optimize the efficiency of radio resources, at the cost of high computational requirements at the radio nodes. Current trends towards Network Function Virtualization aim at enabling software-based networking capabilities at different locations of the radio network architecture. This paper analyzes the impact of moving RRM functions to different Radio Access Network locations and with different channel quality feedback periods. Based on simulation results, the paper analyzes the experienced performance of different channel-aware scheduling strategies in the context of multi-user mobile multimedia scenarios with different radio channel conditions.

Index Terms—CQI feedback delay, CQI reporting rate, wireless networks, scheduling.

I. INTRODUCTION

Historically, the Radio Resource Management (RRM) functions in cellular networks have evolved towards more user-centric schemes. RRM elements are nowadays able to perform enhanced channel-aware scheduling decisions based on real-time feedback provided by User Equipments (UE). This way, cellular networks can meet the objectives of maximizing the efficiency of radio resources while coping with multi-user resource sharing.

Taking 3rd Generation Partnership Project (3GPP) technologies as a reference, the evolution of RRM from 3rd Generation (3G) to 4th Generation (4G) is mainly based on two aspects:

- First, the RRM logic is moved towards the radio interface. In 3G Universal Mobile Telecommunications System (UMTS) technology, the Radio Network Controller (RNC) is the element in charge of RRM and controls a series of Node Bs within the Radio Access Network (RAN). Meanwhile, in 4G Long Term Evolution (LTE) the RRM and other functions are placed at the enhanced Node Bs (eNodeB) with direct access to feedbacks provided by UEs.
- Second, the Transmission Time Interval (TTI) is reduced to support a quicker reaction to instantaneous channel conditions in the multi-user context. In 3G UMTS the possible TTI values are between 10 ms and 80 ms. It was reduced to 2 ms in the transitional High-Speed Downlink Packet Access (HSDPA) technology, and, finally, configured to 1 ms in 4G LTE.

Although the resulting RRM solution maximizes the efficiency of the radio spectrum usage, it faces some issues such as overloading eNodeB elements and associated deployment costs. In future 5th Generation (5G) networks, concepts such as Software Defined Radio (SDR) and Network Function Virtualization (NFV) are being introduced to cope with these challenges. Some RAN-related functions can be moved back to shared hardware elements with high computational capacities. Despite the fact that the low-level RRM functions are likely to remain at the radio edge, some RRM functions can be deployed at the core of the RAN. This possibility can be especially relevant in scenarios where multiple radio accesses need to be coordinated through a unified RRM scheme.

The current tendency for future centralized RRM architectures entails a low-delay reliable mobile fronthaul link to interconnect the centralized radio processing functions and the remote radio head unit. While future fibre fronthaul connections anticipate delays below 1 ms, other types of fronthauls may involve higher delays.

In this context, it is interesting to analyze from a general perspective how far the RRM functions can be moved from the remote radio units in terms of delay. Thus, this paper studies the impact of moving the resource scheduling logic along the RAN with special focus on the associated inaccuracy in channel state information. In order to contextualize the study in real-world deployments, we make use of channel state information obtained from live 4G networks and analyze the associated radio performance variability.

Related works and the specific scope of the paper are analyzed in Section II. Detailed configurations of the analyzed scenario are provided in Section III. This section includes detailed information about the characteristics of traffic patterns and radio channels used in the simulations. Section IV illustrates the simulation results and discusses the obtained RRM performance in different configurations. Finally, Section V provides the conclusions to the paper.

II. BACKGROUND AND SCOPE

In 4G LTE networks, the current performance of the radio channel is reported by UEs by means of the Channel Quality Indicator (CQI) parameter, which is intended to notify the highest supported modulation scheme and coding rate. Regarding CQI reporting, this paper takes into account two significant configuration parameters:

- CQI delay between report generation and its availability at the RRM element. This value is directly related to the duration of the TTI, the location of RRM function and the processing times due to the LTE frame structure.
- CQI Reporting Rate (CRR), as the number of milliseconds from one CQI report to the following one. A low CRR provides better channel-awareness, at the cost of higher signaling overload in the uplink channels.

Several works in the literature aim at analyzing or/and proposing existing or new/enhanced scheduling schemes in the context of partially and inaccurate observable wireless channels (see for example [1]–[7]). Authors in [5], [6] study the impact of the CRR on scheduling with the assumption of no delayed CQI; as concluded, the CRR remarkably affects the performance of channel-aware schedulers. Recently, a low-complexity approach for maximizing the downlink throughput of a system with a partial feedback is proposed in [2]. However, considering simulation results in [4], the system performance is very sensitive to the CQI feedback delay too. In fact, recent work [1] proposes to compute CQIs at user terminal while taking into account the impact of their feedback delay on resource allocation at the base station. Others, such as [3], analyze the effect on scheduling of both the CRR and the CQI feedback delay, but only for a single combination.

To the best of our knowledge, no relevant work exists that rigorously evaluates the combined impact of these two CQI-related components under real-world channel conditions. Therefore, this paper discusses the expected performance in different combinations of CQI delay and CRR. For this aim, a series of simulations were performed including mobile multimedia transmissions and typical scheduling algorithms, i.e. Round Robin (RR), Best CQI (BC) and Proportional Fair (PF). In this case, the three scheduling strategies are somehow channel-aware, in terms of being aware of multi-user CQI values for adaptive modulation and coding. From simulation results, the best performing scheduler can be identified for different CQI configurations and different user mobility patterns.

III. DESCRIPTION OF THE SCENARIO

This paper focuses on the downlink of a 4G network. In LTE, radio resources are divided in radio Resource Blocks (RB), which are assigned to multiple users both in time and frequency [8]. The channel capacity associated to a RB evolves over time due to the intrinsic variability of the medium.

In the last years, the problem of multi-user scheduling in variable channel conditions in heterogeneous contexts has been largely studied. Channel- and user-aware schedulers keep knowledge about the traffic requirements (i.e. per user flow queued bits) and channel state information of the different users, in order to allocate the required resources and also to introduce fairness aspects in the multi-user problem. In time-varying channel conditions, such schedulers take advantage of the channel awareness or opportunistic gains to improve the performance of the system, e.g. the overall achieved throughput. However, reliable instantaneous channel information of each user will not always be available at each decision slot.

A. Traffic pattern

In this paper, traffic demands are based on continuous media sessions; media sessions are assumed to be variable bitrate flows and characterized by its average encoding bitrate. In order to consider a wide range of multimedia services such as radio, music, Standard Definition (SD) video and High Definition (HD) video, media sessions range from 100 kbps up to 1.7 Mbps.

Each media session is composed of a series of media segments, which are transmitted consecutively in time for media continuity. For instance, every Group of Pictures (GOP) in an encoded video results in one different media segment. Assuming variable bitrate flows, media segments of different media sessions are characterized by their size distribution.

Media segments are generated in different media sources and arrive at the RRM element in their path towards the destination mobile user. These media segments are enqueued, and the RRM scheduler decides how many bits are selected to transmit to each user in each TTI.

B. Channel model

Two radio conditions have been analyzed in this paper, considering “static” and “dynamic” users. The performance of radio channels is characterized by means of CQI traces based on experimental tests [9].

Drive tests were performed to gather the experienced performance by UEs in different live LTE networks. Test data were gathered in different mobility patterns and different cities in the North of Spain, using commercial cells in the 2.6 GHz band. Since the CQI traces obtained from the drive test toolset provide a granularity of one second, which is insufficient for the proposed study, we used the Aeroflex 7100 LTE emulator to gather finer-grain radio measurements. The LTE emulator is properly configured to resemble the LTE parameters of the live network. This emulator performs the role of the eNodeB, since it creates the LTE radio signal and all the necessary LTE protocol events to support the attachment and registration of a commercial UE through a radiofrequency cable. The LTE emulator is configured to request CQI reports from the UE every 5 ms, as monitored in live network configurations, and the received CQI values are logged and saved to a file for further offline processing.

In order to emulate the live “dynamic” mobility pattern, the LTE emulator is dynamically configured to change the Signal-to-Noise Ratio (SNR) in a per second basis, following the previously gathered values. Additionally, fading effects are added to the generated LTE signal by means of Pedestrian A model at 7 km/h.

Figure 1 illustrates the emulated SNR conditions and resulting CQI traces obtained from drive testing tool and from the LTE emulator. As can be observed, CQI values are considerably more variable than reported by the field testing equipment. The steady state probabilities of CQI values, denoted as p_1 , are shown in Table I. These state probabilities result in a mean CQI of 8.68.

CQI	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
p_1	0	0	0	0.0012	0.0039	0.0001	0.0308	0.088	0.4431	0.3285	0.0137	0	0	0.0533	0.0307	0.0063
p_{21}	0	0	0	0	0	0	0	0	0	0	0	0.012	0.348	0.520	0.120	0
p_{22}	0	0	0	0	0	0	0	0	0	0.014	0.346	0.520	0.120	0	0	0
p_{23}	0	0	0	0	0	0	0	0.017	0.343	0.520	0.120	0	0	0	0	0
p_{24}	0	0	0	0	0	0.020	0.339	0.520	0.120	0	0	0	0	0	0	0

TABLE I
CQI STEADY STATE DISTRIBUTION FOR DIFFERENT RADIO CONDITIONS.

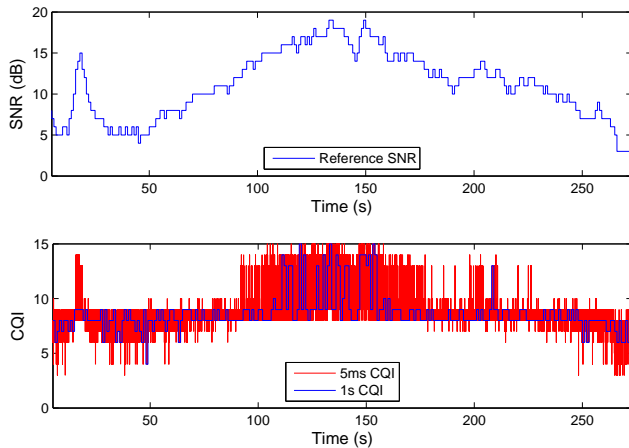


Fig. 1. Radio conditions for "dynamic" users mobility.

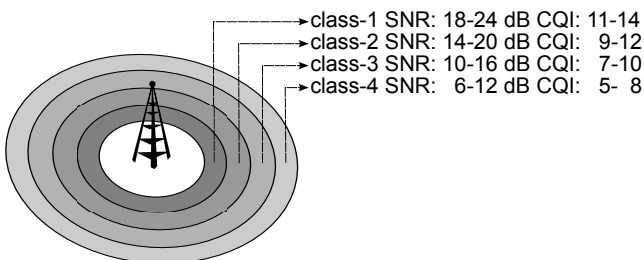


Fig. 2. Different radio conditions used for "static" users mobility.

Similarly for "static" mobility conditions, we performed respective tests in specific radio conditions limiting the SNR variation range for each point of study. Figure 2 illustrates the SNR ranges and the obtained CQI values. As can be observed, the range of CQI values obtained for each condition is more restricted than in the previous "dynamic" case.

Also, obtained CQI traces result in lower temporal variability. The mean CQI burstlength (i.e., number of consecutive CQI samples reported with the same value) is 135 ms, while for the analyzed dynamic setting the CQI burstlength is 26 ms. The per-class steady state distributions of CQIs are provided in Table I; p_{2k} refers to the CQI probability for class k in this second configuration. The mean CQI values from best to worst class are as follows: 12.748, 10.746, 8.743 and 6.703.

C. Schedulers

The scheduler located at the RRM decides the number of RBs to be assigned to each user in each TTI. As previously stated, the resource allocation algorithm uses inaccurate channel-awareness to make decisions. We focus on evaluating the performance of classical schedulers, such as Round Robin (RR), and the popular opportunistic Best CQI (BC) and Proportional Fair (PF) [10]. The advantages of these resource allocation algorithms are their tractability and ease of execution/simulation. Below we provide a brief description of the scheduling policies under study:

- RR: This discipline assigns RBs fairly among users, and for each user transmits the amount of bits according to the latest reported CQI.
- BC: With this policy, the RBs are fairly shared among the users with the estimated highest instantaneous CQI value.
- PF: This policy consists in choosing to transmit the users with the highest ratio of the estimated current transmission rate and the current averaged throughput for a window of 1 s (not taking into account the bits retransmitted); the RBs are assigned fairly among the selected users.

When data from a user media session are transmitted towards the radio channel, if the estimated CQI in the RRM element is higher than the rate available in the transmission moment, we assume that a retransmission needs to happen in the following TTI. In all disciplines, in the case the assigned RB number for each user is higher than its remaining information to transmit, these non-filled RBs are assigned following the priority order established for the different schedulers (randomly in the case of RR). Additionally, a media segment should be downloaded on time for its smooth playout [11]. In this way, media segments that are not fully transmitted before 10 s are discarded [12].

IV. PERFORMANCE EVALUATION

In this section, we analyze the performance of classical schedulers under different imperfect radio channel feedbacks. The performance evaluation is carried out on a simulation-driven environment by taking into account the aspects stated in previous section. In this way, in our controlled experimental platform, we use experimental LTE channel traces obtained from the LTE emulator.

CQI	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r	0	42	67	84	112	168	218	252	268	336	446	504	537	672	756	806

TABLE II
MAPPING OF CQIS TO TRANSMISSION BITS PER ASSIGNED RB.

A. Simulation details

With the purpose of analyzing the joint impact of the CRR and the CQI delay in radio resource allocation, we consider five CRR values (1, 5, 10, 20, 50; in ms) and four CQI delay values (1, 5, 10, 20; in ms) in this study. Note that the actual CQI value at the transmission point is delayed twice the CQI delay respect to the latest reported CQI, due to the uplink and downlink delays.

In reference to the traffic employed in simulations, we consider four types of multimedia sessions for each UE in the cell, with average bitrate of 100 kbps, 300 kbps, 650 kbps and 1.7 Mbps respectively. The media segments of each session arrive according to a Poisson process with a mean rate λ of 2 s. The size distribution of media segments for the aggregate of the four media sessions follows a Pareto distribution with a mean of 170 KB (scale parameter = $2.292 \cdot 10^{-7}$ and shape parameter = 3.974).

As typical in current LTE networks, we use scheduling decision and transmission slots of 1 ms. We assume an LTE cell of 20 MHz, which corresponds to 100 Mbps of theoretical peak downlink rate for UE Cat. 3. Furthermore, we make use of transmission rates associated to the modulation and coding schemes used in LTE networks. Table II provides the mapping between CQI states and bits per RB that can be transmitted by the eNodeB (values extrapolated from [13] for the cell configuration under study).

We have implemented the whole network environment for the scheduling of radio resources in MATLAB. We have carried out simulations of 1000 s length. Both media segments sizes and arrivals are randomly pre-generated; moreover, CQI traces from the LTE emulator are an input vector for the event-driven simulation environment.

We refer to the total network load as the sum of all the loads of video sessions, defined as $\rho = \sum_u \sum_s \lambda \cdot \frac{E[X_s]}{E[r_u]} \cdot TTI$, where $E[X_s]$ is the mean size of a media segment in a video session and $E[r_u]$ the mean user channel rate in bits. In order to study the performance of the scheduling algorithms under severe congestion settings, we have analyzed simulations under high network loads. For that aim, we consider 17 users in the system.

B. Results

In the following, we study the results of different performance metrics for the considered combinations of CQI delays/CRRs and for the three scheduling algorithms employed. First, we analyze the performance in throughput terms. For that purpose, we show the results for the effective throughput, which refers to the rate of received media segments in bits in the simulation time period. Moreover, we

also provide the wasted throughput due to retransmissions. On the other hand, we examine the percentage of received media segments, which is fundamental for multimedia quality assessment. We have analyzed two simulation settings: a single channel setting considering the dynamic channel model provided in Section III, and a multi-class setting employing the static channel model presented in the previous section.

1) *Dynamic setting*: In this first family of simulations all the users in the cell follow the same channel distribution, which corresponds to the dynamic model presented in Section III. If we observe the effective throughput achieved in Figure 3, we conclude that opportunistic policies outperform RR for CQI delay = 1 ms until CRR = 20 ms (except for PF for CRR=20 ms that is notably degraded). However, if we compare the first CQI delay range and CRR=20 ms with the lowest CRR in the next CQI delay value for BC, the decrease of throughput is relevant. Therefore, in order to take advantage of an scheduler aimed at maximizing cell throughput such as BC, the RRM element should be located at 1 ms.

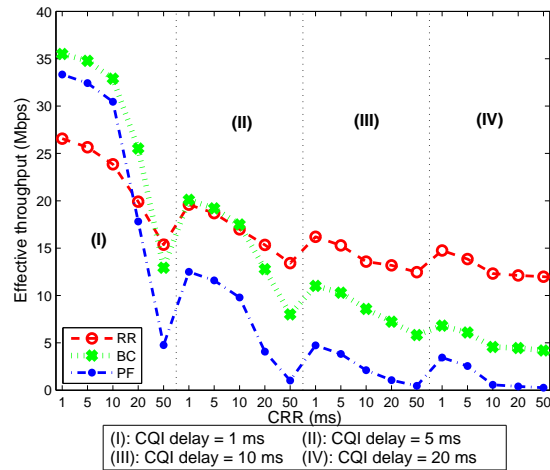


Fig. 3. Effective throughput for the dynamic scenario.

On the other hand, the throughput deterioration for CRR=50 ms is remarkable for the first two CQI delay values. In the second CQI delay range, for CRR={1,5,10}, BC and RR are quite similar, being BC superior and PF the worst. Then, for higher CRRs and higher CQI delays, RR behaves better than the classical channel-aware policies.

Results of the wasted throughput due to retransmissions are collected in Figure 4. It is worth mentioning that BC and PF cause fewer retransmissions than RR in the lowest CQI delay considered until CRR=10 ms. Nevertheless, for the rest of analyzed points the worsening of those two channel-aware policies in this wasted throughput terms is notable.

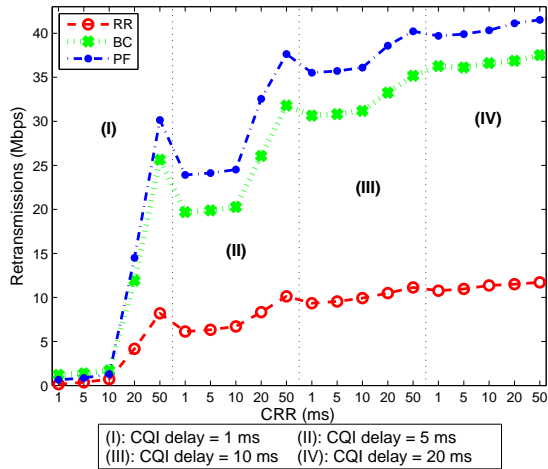


Fig. 4. Wasted throughput caused by retransmission for the dynamic scenario.

As depicted in Figure 5, similarly to the presented behavior for the effective throughput, BC outperforms RR in the percentage of received media segments in the first CQI delay range until the CRR is 20 ms. Nonetheless, RR achieves the highest multimedia quality in the remaining combinations.

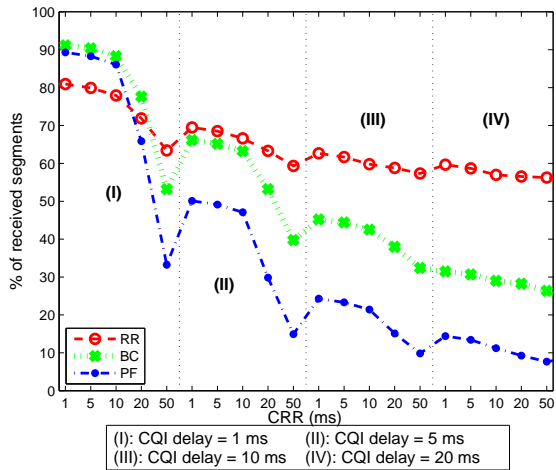


Fig. 5. Percentage of received media segments for the dynamic scenario.

2) *Static setting*: In this scenario we analyze the performance of schedulers under less variable channel models. Furthermore, this is a multi-class setting in CQI terms, where user classes differ due to the UE location in the cell.

Under low mobility conditions, we can see in Figure 6 that BC is the best candidate in effective throughput terms for the whole set of CQI delays/CRRs, which achieves less severe degradations among CQI delay ranges in comparison with the previous scenario. But, as Figure 7 shows, even though the retransmissions are generally lower in this setting, opportunistic schedulers cause more retransmissions than RR.

Besides, Figure 8 summarizes the per-class behavior in received media segments terms for each scheduling discipline. From the plots of the different policies it can be observed that

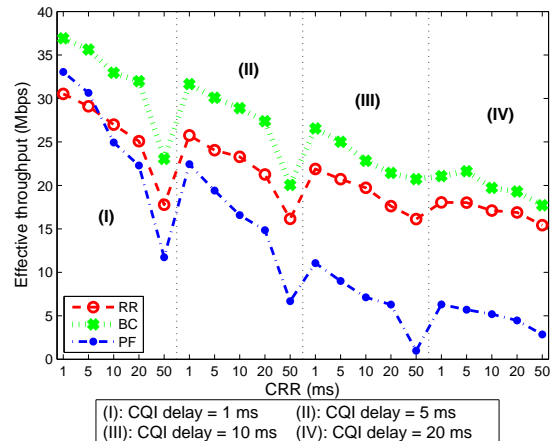


Fig. 6. Effective throughput for the static scenario.

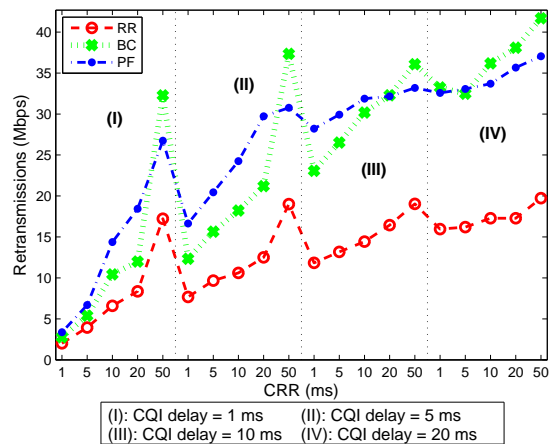


Fig. 7. Wasted throughput caused by retransmissions for the static scenario.

RR and PF are remarkably fairer between classes than BC. However, for BC in the first two CQI delay ranges, except for CRR=50 ms, the percentages of received media segments for the best two classes are extraordinary (for the best class in all the analyzed points). Therefore, when choosing among schedulers, it would depend on the degree of fairness it is expected to achieve.

V. CONCLUSIONS

This paper focuses on the study of the impact of two significant parameters related to channel quality feedbacks sent by mobile UEs to the RRM element in mobile broadband networks.

First, the CQI delay is the time lapse since the channel quality measurement until the downlink Radio Resource Control (RRC) decision is implemented, and it is imposed by the radio technology TTI and the location of the RRC within the RAN. Second, the CRR determines the frequency of channel feedbacks, and its value is a typical trade-off between accuracy in the channel-awareness and uplink signaling overload.

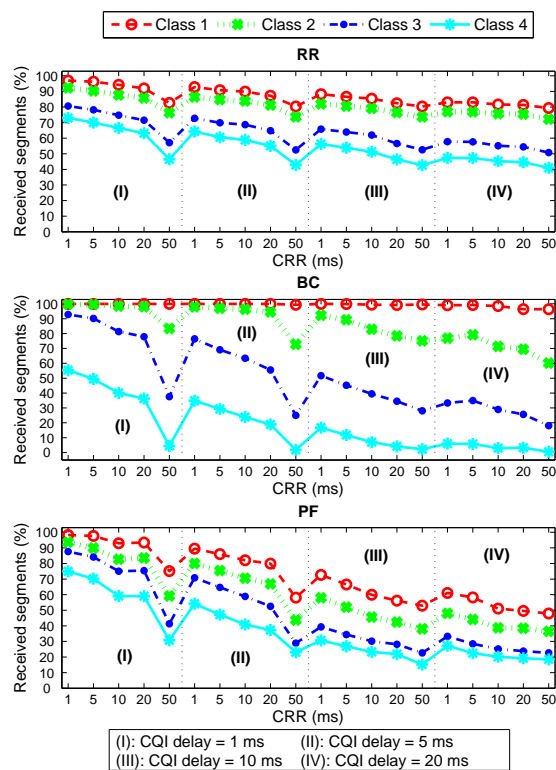


Fig. 8. Percentage of received media segments for the static scenario.

Through simulations, we compare the experienced performance in different combinations of these two values under the governance of three classical scheduling strategies: RR, BC and PF. The performance of the radio cell is evaluated in terms of total throughput, required radio retransmissions and percentage of received media segments. These parameters have been evaluated in two scenarios: highly mobile users and more static multi-class users. The present study has three main differential aspects compared to background studies:

- We use CQI traces based on experimental measurements rather than simulation or model-based channel estimations. Although the scope of the results may be limited to the experimental mobility patterns, they are more credible than non-realistic channel traces.
- In our case, all schedulers are user- and channel-aware in the sense that they assign the amount of bits to RBs based on the latest received CQI. Thus, even if RR assigns more RBs than needed by the instantaneous traffic demand for a user, only the required RBs are used and the rest of resources are shared to other users.
- We focus on traffic patterns associated to multimedia sessions, allowing discards of media segments, rather than the typical greedy sources. This aspect entails that, in BC, users in best conditions do not consume all the radio resources.

The discussed results allow evaluating the convenience of selecting different types of scheduling strategies in function

of the configurable parameters CQI delay and CRR. For example, in high mobility scenarios, BC achieves higher average throughput for up to CRR=20 ms when the RRM logic is at the eNodeB. Meanwhile, it only achieves higher throughput up to CRR=10 ms for CQI delays of 5 ms. For the rest of configurations, the high variability of CQI values increases the retransmissions due to inaccurate CQI estimations and RR performs better. With low mobility multi-class users, BC always behaves better since the CQI variability is reduced and the required radio retransmissions decrease.

ACKNOWLEDGMENTS

This work has been partially funded by the Spanish Ministerio de Economía y Competitividad (MINECO) under grant TEC2013-46766-R: QoEverage - QoE-aware optimization mechanisms for next generation networks and services.

REFERENCES

- [1] Ayaz Ahmad, Naveed Ul Hassan, and Nadir Shah, "Robust channel quality indicator reporting for multi-carrier and multi-user systems," *Computer Networks*, vol. 74, pp. 78–88, 2014.
- [2] Yingxiang Yang, Bo Bai, Wei Chen, and Lajos Hanzo, "A Low-complexity Cross-Layer Algorithm for Down-Link Coordinated Scheduling and Robust Beamforming Under a Limited Feedback Constraint," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, 2014.
- [3] R Basukala, HAM Ramli, K Sandrasegaran, and Lu Chen, "Impact of CQI feedback rate/delay on scheduling video streaming services in LTE downlink," in *Communication Technology (ICCT), 2010 12th IEEE International Conference on*. IEEE, 2010, pp. 1349–1352.
- [4] Ayaz Ahmad and Mohamad Assaad, "Optimal power and subcarriers allocation in downlink OFDMA system with imperfect channel knowledge," *Optimization and Engineering*, vol. 14, no. 3, pp. 477–499, 2013.
- [5] Ianire Taboada, Jose Oscar Fajardo, Fidel Liberal, and Bego Blanco, "Size-based and channel-aware scheduling algorithm proposal for mean delay optimization in wireless networks," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 6596–6600.
- [6] Ianire Taboada, Jose Oscar Fajardo, and Fidel Liberal, "Performance Analysis of Scheduling Algorithms for Web QoE Optimization in Wireless Networks," *Network Protocols and Algorithms*, vol. 4, no. 4, pp. 27–43, 2012.
- [7] Stefan Schwarz, Robert W Heath Jr, and Markus Rupp, "Single-user MIMO versus multi-user MIMO in distributed antenna systems with limited feedback," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–20, 2013.
- [8] F. Capozzi, G. Piro, L.A. Grieco, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.
- [9] J. O. Fajardo, I. Taboada, and L. Liberal, "Analysis of cqi traces from lte mimo deployments and impact on classical schedulers," in *13th International Conference on Wired and Wireless Internet Communications (WWIC)*, 2015, pp. 59–72.
- [10] S. Aalto and P. Lassila, "Flow-level stability and performance of channel-aware priority-based schedulers," in *Next Generation Internet (NGI), 2010 6th EURO-NF Conference on*. IEEE, 2010, pp. 1–8.
- [11] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial Delay vs. Interruptions: Between the Devil and the Deep Blue Sea," in *Proceedings of the 2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2012, pp. 1–6.
- [12] 3GPP, "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)," *3GPP TS 26.247*, 2014.
- [13] S. Sesia, I. Toufik, and M. Baker, *LTE-The UMTS Long Term Evolution: From Theory to Practice*, Wiley, 2011.

REMODE: Un algoritmo para el cálculo de la reputación de nodos y caminos en redes ad hoc cooperativas

Lenin Guaya Delgado, Esteve Pallarès Segarra, Jordi Forné
Departament d'Enginyeria Telemàtica.

Universitat Politècnica de Catalunya (UPC).

Barcelona, España. CP: 08034
{lenin.guaya, esteve, jforne}@entel.upc.edu

Resumen- En este artículo se presenta REMODE (Reputation for Mobile Devices), un algoritmo para estimar la reputación de reenvío de los nodos de una red MANET. El objetivo final es poder elegir aquellos caminos que estén formados por los nodos con mayor reputación, evitando así los nodos menos colaborativos. La estimación de la reputación se realiza utilizando solamente información de primera mano que dispone el nodo en base a su experiencia en la red. Se presenta un análisis teórico sobre los valores estimados de reputación. Además el algoritmo se ha implementado en el simulador NS-2.35, y se han ejecutado simulaciones en escenario fijo y móvil, el escenario móvil ha sido creado con Bonnmotion v2.1a. Los resultados de dicho algoritmo dependen del comportamiento de los nodos que participan en cada camino. Finalmente se presentan los resultados comparando los valores estimados de reputación (tanto teóricos como simulados), con los valores reales de reputación que debería corresponder a los nodos.

Palabras Clave- Reputación, MANET, encaminamiento, comportamiento egoísta.

I. INTRODUCCIÓN

Reputación es la opinión de una entidad acerca de otra [1]. Reputación es el conocimiento del pasado, una memoria vinculada a una entidad específica, por tanto al agregar el concepto de reputación a redes móviles ad hoc (MANETs) se puede predecir el comportamiento de los nodos que conforman la red.

Un conjunto de dispositivos móviles inalámbricos con capacidades de enrutamiento forman una MANET, una red con enrutamiento distribuido y topología cambiante. Al ser el enrutamiento distribuido el algoritmo de reputación debe dar información suficiente a cada nodo para permitirle tomar sus propias decisiones de enrutamiento, y al ser la topología cambiante, el algoritmo debe mantener un valor estable a pesar de las constantes rupturas de los enlaces. Tomando éstas consideraciones previas, la presente investigación presenta un nuevo algoritmo que tiene la capacidad de asignar a cada nodo un valor probabilístico de reputación calculado en base a su participación en los caminos, estos valores son bastante estables debido a la presencia de dos ecuaciones principales que usan media móvil.

En MANETs existen dos tipos de mal comportamiento, egoísta y malicioso. En el egoísta el nodo usa la red pero no coopera, simplemente ahorra energía, memoria o ciclos de CPU para sus propias comunicaciones. En el malicioso el nodo daña a otros nodos haciendo caer la red, ahorrar energía

no es su prioridad [2]. Este trabajo se enfoca en el mal comportamiento egoísta, y se simula este comportamiento con ayuda de la implementación de un modelo de pérdidas aleatorio sobre el protocolo DSR del simulador NS-2.

En REMODE los nodos observan la reputación de los caminos, y a partir de este muestreo pueden inferir la reputación de: (i) los nodos, (ii) otros caminos posibles que no han sido rutas activas pero que están en caché del nodo origen. Esto es similar al algoritmo RDSR-V [3].

La innovación de la propuesta es que la reputación se asocia directamente a la probabilidad de no enrutamiento por parte del nodo o camino, además el cálculo es mejor que en RDSR-V [3].

El resto del artículo se estructura de la siguiente forma. En la sección II se realiza una breve descripción de redes móviles ad hoc y del protocolo de enrutamiento DSR, a la vez que se presenta el estado del arte en sistemas de reputación en MANETs. La sección III presenta el algoritmo propuesto, detallando las ecuaciones de cálculo de la reputación para nodos y caminos. En la sección IV se evalúa el funcionamiento del algoritmo en tres escenarios estáticos y uno móvil. Finalmente, en la sección V se presentan las conclusiones.

II. ESTADO DEL ARTE.

A. MANETs.

El conjunto de terminales de datos digitales con transceptores inalámbricos sin infraestructura de red y sin control centralizado forman una MANET. En una MANET, cuando un nodo quiere transmitir un paquete a otro nodo que está localizado fuera del alcance, el paquete pasa a través de una secuencia de nodos intermedios usando un principio de transmisión de almacenamiento y reenvío "multisalto". Las MANETs son usadas en redes de comunicaciones militares, en campos de batalla, operaciones de rescate de emergencia, operaciones submarinas, monitoreo ambiental y exploración espacial, que son entornos en los cuales es necesario un despliegue de red "sobre la marcha" [4].

B. Descripción de DSR.

Dynamic Source Routing (DSR) es un protocolo simple y eficiente diseñado específicamente para uso en redes inalámbricas multisalto. DSR da a la red la capacidad de

autoorganizarse y autoconfigurarse sin el uso de infraestructura o administración centralizada. DSR trabaja con muy altas tasas de movilidad y hasta 200 nodos.

En DSR se identifica dos principales mecanismos, el “Descubrimiento de Ruta” y el “Mantenimiento de Ruta”, para descubrir y mantener caminos en la red ad hoc [5].

DSR es un protocolo reactivo, esto quiere decir que busca un camino únicamente cuando lo necesita, en contraparte a los protocolos proactivos que buscan caminos periódicamente.

C. Sistemas de Reputación en MANETs.

Los sistemas de reputación son una de las mejores soluciones para tratar con el mal comportamiento egoísta, por tanto es pertinente comentar brevemente los objetivos y componentes de estos sistemas.

El objetivo de los sistemas basados en reputación es proveer información para distinguir nodos de buena reputación de nodos de mala reputación.

Un sistema de reputación tiene cuatro componentes, el primero es *recolección de información* que se encarga de recolectar información acerca de los nodos, el segundo es *compartición de información* responsable de la diseminación de la información de primera mano reunida por los nodos, el tercero es *modelado de la información* que combina información de primera mano con información de segunda mano en una métrica, también es responsable de mantener y actualizar esta métrica; y, finalmente el cuarto componente es *toma de decisiones* encargado de hacer todas las decisiones basándose en información proveída por el componente de *modelado de la información* [6].

D. Descripción de otros algoritmos de protocolos basados en reputación.

RDSR-V [3], OCEAN [7], CONFIDANT [8] y CORE [9] son implementados sobre DSR, y cada entidad de red mantiene un histórico de lo colaborativas que son otras entidades mediante una técnica de reputación que mejora la cooperación de los nodos y por tanto el rendimiento de la red. Estos cuatro algoritmos son descritos a continuación:

El algoritmo que forma la métrica de reputación de RDSR-V usa dos medias móviles para calcular valores probabilísticos de reputación. En RDSR-V el peso dado a alfa para nuevos valores es dividido para el número de nodos del camino, y la función de muestreo para el cálculo de la reputación muestreada es una exponencial $e^{-14xFPLp}$.

El algoritmo que forma la métrica de reputación de OCEAN, usa información de primera mano, es decir observaciones directas, por tanto es menos vulnerable a falsas acusaciones. En este algoritmo, se nota la presencia del componente Route Ranker, el cuál con eventos positivos suma uno, con eventos negativos resta dos, y cuando alcanza un umbral de -40 el nodo es registrado en una lista de fallo.

El algoritmo que forma la métrica de reputación de CONFIDANT se compone de tres métricas, la primera es el *goodput* G que es un valor porcentual de 0 a 100 obtenido de la relación entre paquetes recibidos y paquetes originados, la segunda métrica es el *overhead* O el cual expresa la sobrecarga causada por mensajes extra; y, la tercera métrica es *utility* U_i la cual indica el costo-beneficio de un nodo.

Finalmente, se tiene que, el algoritmo que forma la métrica de reputación de CORE se compone de tres métricas, la *reputación subjetiva* que es calculada a partir de

información de primera mano, la segunda métrica es la *reputación indirecta* que es actualizada desde información de segunda mano; y, la tercera métrica es la *Reputación Funcional* que es la combinación de reputación subjetiva e indirecta para diferentes funciones. Además, se nota que los valores de reputación de CORE varían entre -1 y +1.

III. ALGORITMO REMODE

El objetivo de este algoritmo es elegir, de entre los caminos disponibles, aquellos que tengan una mayor reputación, o lo que es lo mismo, los caminos que tengan menos pérdidas de paquetes. Para ello, cada nodo origen estima la reputación de sus nodos vecinos a partir de las pérdidas que se producen en los caminos que utiliza. La estimación de la reputación de los nodos se realiza repartiendo las pérdidas de cada camino por igual entre los nodos intermedios de dicho camino. Ello, aunque pueda parecer injusto para los nodos, no lo es, ya que el objetivo del algoritmo no es estimar dicha reputación con exactitud, si no conocer los caminos con menos pérdidas. Cuando un nodo origen dispone de más de un camino para alcanzar un destino, calcula la reputación de dichos caminos a partir de las estimación de la reputación de los nodos, eligiendo aquel camino que presente menores pérdidas, es decir, mayor reputación.

La Fig. 1 representa los nodos que pertenecen a un camino. Si alguno de estos nodos tiene un comportamiento egoísta, no reenviará una fracción de los paquetes recibidos. Se define la fracción de pérdida de paquetes por el nodo *j* como la relación entre el número de paquetes no reenviados y la cantidad total de paquetes recibidos (*FPLn_j* -fraction of packet losses - [3]). Para estimar el valor de reputación de reenvío de un nodo *j* (*RFVn_j* -reputation forwarding value- [3]), se supone que el principal motivo por el cual se pierden los paquetes es el comportamiento egoísta de determinados nodos de la red. Se define la reputación de reenvío para un nodo *j* como la probabilidad complementaria de la fracción de pérdida de paquetes (Ec. 1).

$$RFVn_j = 1 - FPLn_j \quad (1)$$

En esta propuesta se asume que el nodo origen (Fig. 1) recibe información periódica sobre la fracción de pérdida de paquetes para cada camino *i* (*FPLp_i*) que esté utilizando. Dicha periodicidad puede depender de la cantidad de tráfico cursado en dicho camino, o bien puede ser cada cierto intervalo de tiempo fijo. Queda fuera de los objetivos de este trabajo definir la forma en la que los nodos reciben esta información. El valor complementario de la *FPLp_i* es la reputación de reenvío asociada a dicho camino (*RFVp_i*). El emisor promedia dicho valor mediante una media móvil (Ec. 2)

$$\overline{RFVp_i}|_{updated} = (1 - \alpha) \overline{RFVp_i}|_{old} + \alpha (1 - FPLp_i) \quad (2)$$

Por otro lado, considerando que el comportamiento de cada nodo a la hora de realizar los reenvíos es independiente del resto, se puede calcular analíticamente la reputación de un camino *i* (*RFVp_i*) como un producto de probabilidades (Ec. 3). En la siguiente sección a esta reputación se la denomina SCORE, ya que nos indica la calificación del camino.

$$RFVp_i = \prod_{j \in path_i} (RFVn_j) \quad (3)$$

Dado que a priori el nodo origen desconoce el comportamiento de los nodos que forman un camino, no es

posible utilizar la Ec. 3 para estimar la reputación de reenvío de dichos nodos. El algoritmo propuesto para realizar dicha estimación, se basa en repartir las pérdidas por igual entre los nodos que forman el camino, excepto el nodo origen que se supone que no genera pérdidas. En este caso, se puede reescribir la Ec. 3 de la siguiente forma.

$$RFVp_i = (RFVn_j)^{n_i} \quad (4)$$

Siendo n_i el número de nodos intermedios en el camino i . Cada vez que el nodo emisor actualiza la fracción de pérdida de paquetes de un camino i , se estima el valor de la reputación de reenvío de cada uno de los nodos que pertenecen a dicho camino ($\overline{RFVn_j}$). Para ello se utiliza el valor de la reputación del camino i calculada en Ec. 2 sobre la Ec. 4.

$$\overline{RFVn_j} = (\overline{RFVp_i})^{\frac{1}{n_i}} \quad (5)$$

El hecho de repartir las pérdidas por igual entre los nodos que componen el camino perjudica a la reputación de los nodos que tienen un comportamiento colaborativo y beneficia a la reputación de los nodos con un comportamiento más egoísta. Es por ello que la reputación de los nodos debe promediarse teniendo en cuenta las estimaciones hechas en todos los caminos en los que el nodo participa. Para este caso también se realiza una media móvil (Ec. 6).

$$\overline{RFVn_j}|_{\text{updated}} = (1 - \beta)\overline{RFVn_j}|_{\text{old}} + \beta \overline{RFVn_j} \quad (6)$$

El objetivo final es que cada nodo disponga de una tabla con las reputaciones de los otros nodos de la red, que le permita evaluar la reputación de los distintos caminos que ha descubierto y elegir aquel de mayor reputación. La reputación de los caminos se obtiene utilizando de nuevo la Ec. 3.

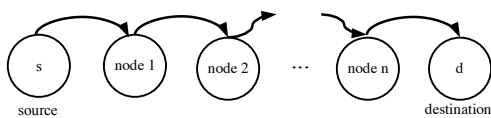


Fig. 1: Nodos que componen un camino.

Es posible realizar un análisis teórico para obtener el valor medio de la reputación que un nodo origen s , asignará a un nodo intermedio j mediante el algoritmo propuesto. Para el caso que el nodo origen reciba información sobre las pérdidas de los caminos cada vez que manda un determinado volumen de tráfico el cálculo dependerá del número de caminos establecidos y el tráfico cursado en cada uno de ellos, tal como se muestra en la Ec. 7.

$$\overline{RFVn_{j,s}} = \sum_{i \in \{n_s, n_j\}} \left(\frac{\gamma_i}{\gamma_{n_{j,s}}} (RFVp_i)^{\frac{1}{n_i}} \right) \quad (7)$$

Donde:

- i : Caminos con origen en el nodo s , a los cuáles pertenece el nodo j .

n_i : Número de nodos intermedios del camino i .

γ_i : Tráfico cursado en el camino i .

$\gamma_{n_{j,s}}$: Tráfico de todos los caminos con origen en el nodo s , a los cuáles pertenece el nodo j (Ec. 8).

$$\gamma_{n_{j,s}} = \sum_{i \in \{n_s, n_j\}} (\gamma_i) \quad (8)$$

Para el caso que el nodo origen reciba información sobre las pérdidas cada cierto tiempo fijo, todos los caminos tendrán el mismo peso independientemente del tráfico cursado. Ello equivale a utilizar la Ec. 7 con todos los tráficos iguales.

En la Ec. 7 la $RFVp_i$ de los caminos se obtiene a partir de la Ec. 3, tomando como dato la fracción de pérdida de paquetes en cada nodo (Ec.2).

Así por ejemplo, en la Fig. 2 se representa una topología con distintos caminos, siendo los nodos 7 y 9 los únicos que no retransmiten todos los paquetes que reciben (con pérdidas). El nodo 5 es el nodo origen para los caminos 5, 6 y 7. Tomando el mismo tráfico en todos los caminos se obtiene que la reputación que el nodo 5 le asignará al nodo 6 es 0,9426 (Ec. 9), cuando en realidad le correspondería una reputación de 1 ya que reenvía todos los paquetes.

$$\overline{RFVn_{6,5}} = \frac{1}{3} \cdot 1 + \frac{1}{3} (0,75 \cdot 0,85)^{\frac{1}{5}} + \frac{1}{3} (0,75 \cdot 0,85)^{\frac{1}{5}} = 0,9426 \quad (9)$$

IV. RESULTADOS Y ANÁLISIS.

Se ha implementado el algoritmo en el simulador NS-2 [10], y se han realizado distintas simulaciones tanto en escenarios fijos como móviles.

El primer escenario planteado es el de la Fig. 2 con los parámetros de la Tabla I, consta de 17 nodos dispuestos de tal forma que nos dan un total de 8 caminos, en color rojo se muestran los caminos no cooperativos (2, 4, 5 y 6) y en azul el resto (1, 3, 7 y 8). Dado que el protocolo DSR debe realizar un *discovery* para encontrar los posibles caminos hacia los distintos destinos, se han ubicado los nodos estratégicamente para forzar los caminos que aparecen en la Fig. 2.

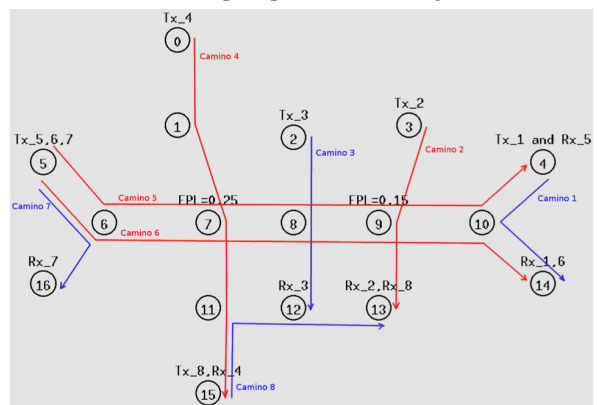


Fig. 2. Diseño del escenario de simulación.

Lo primero que se analizará es lo que sucede cuando se varían los valores de los pesos α y β (Ec. 2 y Ec. 6)

Inicialmente se varía el valor de α . En la Fig. 3 se observa las reputaciones de 5 caminos, 4 que incluyen nodos no cooperativos (caminos 2, 4, 5 y 6) y uno con nodos cooperativos (camino 3), tanto para $\alpha=0,25$ (línea con signos +) como para $\alpha=0,05$ (línea continua). Observamos claramente que con $\alpha=0,05$ los valores de \overline{RFVp}_i tardan más en actualizarse, es decir presentan un transitorio mayor. Además observamos que el valor de los distintos transitorios no depende significativamente del valor final de la reputación del camino, sino que depende únicamente del valor de α . Para el caso del camino 5 con $\alpha=0,25$, dicho transitorio se inicia más tarde debido a que el nodo origen ha necesitado más tiempo para realizar el *discovery*. Transcurrido el transitorio los valores de las reputaciones se estabilizan alrededor del valor que se obtiene a partir de la Ec. 3, el cual coincide para los caminos 5 y 6 como se deduce de la Fig. 2. Cuanto mayor es α , mayor es la variación alrededor de dicho valor, de manera que existe un compromiso entre la duración del transitorio y la varianza alrededor de un valor estable de reputación.

Tabla I
PARÁMETROS DE LA SIMULACIÓN

Parámetro	Valor
Simulador	ns-2 v2.35
Área	410 m x 290 m
Número de nodos	17
Número de conexiones CBR	8
Rango de transmisión.	95 m
Especificación MAC.	IEEE 802.11g
Patrón de movimiento.	Estático.
Ancho de banda nominal.	54 Mbps
Tiempo de simulación	700 s
Protocolo de enrutamiento.	DSR con nodos egoístas.
Protocolo de transporte.	UDP
Tipo de tráfico	CBR sobre UDP
Máximo tamaño de paquete.	1500 Bytes
Intervalo de tráfico CBR.	0.005
Número de fuentes.	6
Nodos fuente.	Nodos 4,3,2,0,5,15
Número de destinos.	6
Nodos destino.	Nodos 14, 13, 12,15, 4 y 16
Tamaño de cola.	50 paquetes
Número de nodos egoístas	2
Nodos que generan las pérdidas	Nodo 7 y nodo 9
Tiempo de muestreo.	Cada 5 s.
Intervalo de tiempo de transmisión.	0 s.- 700 s.

Los valores de \overline{RFVp}_i de los caminos 1, 7 y 8 siempre son iguales a 1, por ello no constan, en su lugar sólo se grafica la reputación del $RFVp_3$, que también es igual a 1.

En un segundo análisis se observa lo que sucede cuando se varía β , que es un peso que afecta directamente a los nodos (Ec. 6).

En las Fig. 4 y 5 se muestra la reputación del nodo 6 (\overline{RFVn}_6) calculada por el nodo 5, ya que el nodo 5 es el nodo origen de los caminos en los que participa el nodo 6 (caminos 5, 6 y 7). La reputación que finalmente el nodo 5 asigna al nodo 6 es inferior a su reputación real (que vale 1 al ser el nodo 6 un nodo cooperativo) debido a que en los caminos en que participa el nodo 6 existen nodos con un comportamiento egoísta.

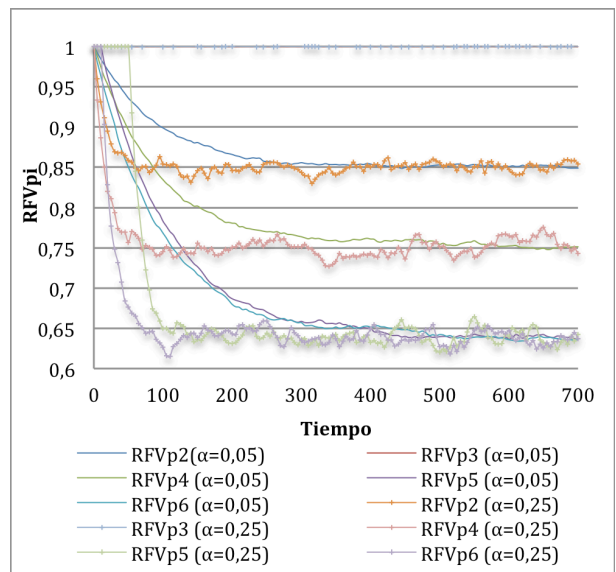


Fig. 3. \overline{RFVp}_i variando α con $\beta=0,05$.

En la Fig. 4 se muestran los valores de \overline{RFVn}_6 para $\beta=0,05$ y $\beta=0,25$ con $\alpha=0,25$ constante. Al igual que ocurre con la reputación de los caminos, se nota mayor varianza al aumentar β , así para $\beta=0,05$ la varianza es de $5,260 \times 10^{-7}$, mientras que para $\beta=0,25$ la varianza es de $6,612 \times 10^{-6}$. También se muestra que para un mayor valor de β el transitorio es menor.

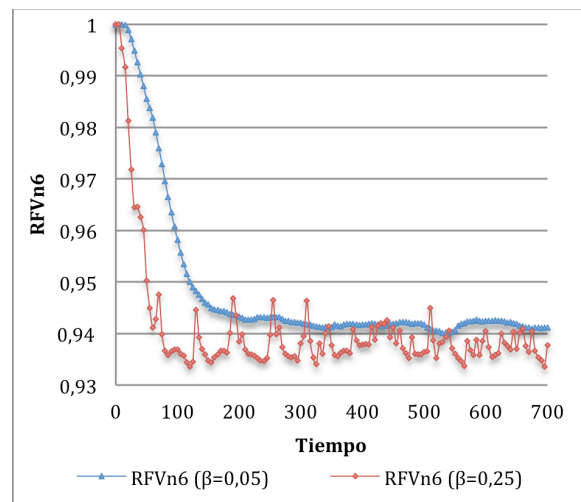


Fig. 4. \overline{RFVn}_6 variando β con $\alpha=0,25$.

En la Fig. 5 se muestran los mismos resultados para un valor de $\alpha=0,05$. Para $\beta=0,05$ la varianza es de $1,0025 \times 10^{-5}$, y para $\beta=0,25$ la varianza es de $1,6483 \times 10^{-5}$. En este caso un mayor valor de β también implica un aumento de la varianza, pero no una mejora en el transitorio. Ello es debido a que para un valor de α tan pequeño la actualización de los caminos es muy lenta y enmascara el efecto que pueda tener β sobre el transitorio de la reputación de los nodos.

Ahora planteamos un segundo escenario con la misma topología de la Fig. 2, pero aumentando las pérdidas de los nodos no colaborativos ($FPLn_7=0,4$ y $FPLn_9=0,3$) y usando $\alpha=\beta=0,25$.

La Fig.6 muestra la reputación muestreada de los caminos, esta es la información que utilizan los nodos origen de cada

camino para calcular la \overline{RFVp}_i , luego la \overline{RFVn}_j y finalmente obtienen las calificaciones de cada camino.

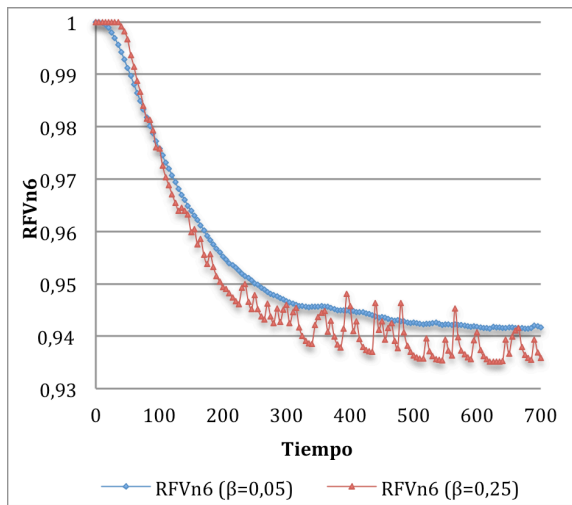


Fig. 5. \overline{RFVn}_6 variando β con $\alpha=0,05$

La reputación de los caminos 1, 7, 8 al igual que el camino 3 son iguales a 1, por ello no se los ha representado, facilitando así, la lectura de la gráfica.

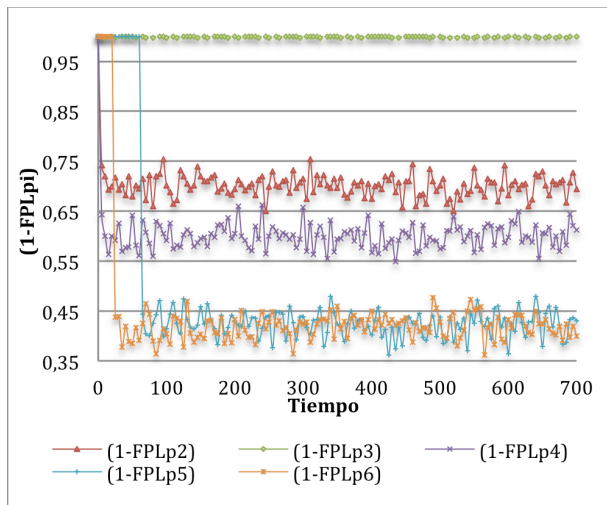


Fig. 6. Complemento de la FPL muestreada por cada camino.

En la Tabla II se muestra cuál es el cálculo de los valores de reputación de los nodos vistos desde cada nodo origen. El nodo origen 4 únicamente muestrea información del camino 1, el nodo 3 del camino 2, el nodo 2 del camino 3, el nodo 0 del camino 4, el nodo 5 de los caminos 5, 6 y 7, y finalmente para el nodo 15 sólo está a su alcance muestrear el camino 8. Por tanto el nodo origen 4 únicamente calculará la \overline{RFVn}_j del nodo intermedio 10, el nodo origen 3 del nodo intermedio 9, el nodo origen 2 de nodo intermedio 8, el nodo origen 0 de los nodos intermedios 1, 7 y 11, el nodo origen 5 de los nodos intermedios 6, 7, 8, 9 y 10, el nodo origen 15 de los nodos intermedios 11 y 12.

Además, la Tabla II muestra una comparativa entre los valores reales, teóricos y los calculados a partir de la simulación, en donde se nota que los valores teóricos y simulados son prácticamente iguales, verificando la utilidad de la Ec. 7. Los valores simulados que se han tabulado

corresponden al valor de reputación calculado en el último instante de la simulación. La diferencia entre el valor teórico y el final simulado es debido a la varianza que presenta el método propuesto, y que el resultado de la Ec. 7 es un valor medio. Adicionalmente se muestran los intervalos de confianza de los valores simulados promedio (de 210 a 700 s.).

Tabla II
REPUTACIÓN DE NODOS OBSERVADA POR EL NODO ORIGEN CUANDO SE INCREMENTA LA FPL.

Nodo	Val. Real	Val. Teór.	Val. Sim. (promedio)	Int. Conf. 95% de Val. Sim. (promedio)
N_0				
N_1	1	0,8434	0,8435	(0,8434; 0,8445)
N_7	0,6	0,8434	0,8435	(0,8434; 0,8445)
N_{11}	1	0,8434	0,8435	(0,8434; 0,8445)
N_2				
N_8	1	1	1	(1,0000; 1,0000)
N_3				
N_9	0,7	0,7	0,6996	(0,6996; 0,7017)
N_4				
N_{10}	1	1	1	(1,0000; 1,0000)
N_5				
N_6	1	0,8938	0,8817	(0,8807; 0,8887)
N_7	0,6	0,8407	0,8414	(0,8390; 0,8481)
N_8	1	0,8407	0,8414	(0,8390; 0,8481)
N_9	0,7	0,8407	0,8414	(0,8390; 0,8481)
N_{10}	1	0,8407	0,8414	(0,8390; 0,8481)
N_{15}				
N_{11}	1	1	1	(1,0000; 1,0000)
N_{12}	1	1	1	(1,0000; 1,0000)

En la Tabla II se observa que los nodos 2 y 3 obtienen la reputación real de los nodos 8 y 9 respectivamente, ya que éstos son los únicos nodos intermedios para los caminos 3 y 2 respectivamente. Por el contrario el nodo 5 asigna la misma reputación a ambos nodos, lo cual beneficia a la reputación real del nodo 9 (el egoísta) en detrimento de la reputación de nodo 8 (colaborativo). El nodo 5 asigna la misma reputación a todos los nodos excepto al nodo 6, debido a que éste participa en el camino 7. En este caso el nodo 6 no está afectado por el comportamiento malicioso de los nodos 7 y 9, siendo su reputación más alta que la del resto de nodos.

Tabla III
CALIFICACIONES DE CAMINOS OBSERVADA POR EL NODO ORIGEN CUANDO SE INCREMENTA LA FPL.

Nodo	Cal. Real	Cal. Sim (700 s.)	Cal. Real-Cal. Sim.
N_0			
$C_{1-3, 5-8}$	1	1	0
C_4	0,6	0,6008	-0,0008
N_2			
C_{1-8}	1	1	0
N_3			
$C_{1, 3-8}$	1	1	0
C_2	0,7	0,7026	-0,0026
N_4			
C_{1-8}	1	1	0
N_5			
$C_{1-4, 8}$	1	1	0
C_5	0,42	0,4369	-0,0169
C_6	0,42	0,4415	-0,0215
C_7	1	0,9108	0,0892
N_{15}			
C_{1-8}	1	1	0

Cuando un nodo tiene paquetes para transmitir primeramente busca en su caché si tiene algún camino al destino solicitado, en caso de no tenerlo inicia un *discovery*

para obtener el camino. Ahora partiendo de que se tiene una o más rutas en caché, con el algoritmo aquí presentado se pueden obtener las calificaciones para cada camino, y en base a ellas seleccionar el camino de mayor reputación al destino solicitado. En nuestro caso, para la topología de la Fig. 2, las calificaciones para cada camino, obtenidas a partir de las reputaciones de los nodos calculadas por cada nodo origen, son las que se muestran en la Tabla III. Se observa que calificación real y simulada tienen una buena aproximación, a pesar que el objetivo del algoritmo no es obtener el valor exacto, si no discriminar cual es el camino con mayor reputación.

En los escenarios anteriores cada nodo origen disponía de una única ruta activa, debido a que las distancias entre nodos fueron dispuestas de tal forma que se formen un único camino para cada destino. En este nuevo experimento se plantea la posibilidad que un nodo origen disponga de más de una ruta hacia el mismo destino. Para ello se ha utilizado la topología de la Fig. 2, tomando el nodo 8 como único origen, y los nodos 0, 3, 4, 14, 13, 15, 16 y 5 sus destinos, formando los caminos 1, 2, 3, 4, 5, 6, 7 y 8 respectivamente, tal como se muestra en Tabla IV. Para la simulación se han utilizado los parámetros de la Tabla I considerando $\alpha=\beta=0,25$, y pérdidas $FPLn_7=0,25$ y $FPLn_9=0,15$.

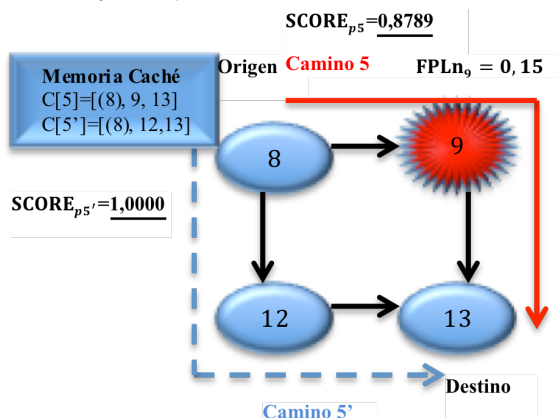


Fig. 7. Calificaciones para caminos 5 y 5' calculadas por REMODE.

En este escenario el nodo 8, para transmitir datos al 13, tiene dos caminos disponibles tal como lo muestra la Fig. 7. Ambas rutas tienen dos saltos para llegar al nodo 13. El protocolo DSR elige el primero de los caminos que descubre, en cambio el protocolo REMODE compararía ambos caminos una vez descubiertos y compararía sus reputaciones para tomar la decisión. En la simulación la calificación para el camino 5 es 0,8789, mientras que para el camino 5' es de 1,0000, siendo este último la mejor opción.

Finalmente, se presenta un cuarto escenario, usando los nodos y caminos definidos en la Fig. 2, añadiendo a los nodos un patrón de movimiento *Random Waypoint* [11] con una rapidez de los nodos de 2m/s. Los parámetros de la simulación son los definidos en la Tabla I, con $\alpha=\beta=0,25$, así como $FPLn_7=0,25$ y $FPLn_9=0,15$. La posición inicial de los nodos es aleatoria y es la que se muestra en la Fig. 8.

La Fig. 9 muestra la reputación cambiante que presentan los caminos. Durante la simulación cada camino va cambiando de nodos intermedios, manteniendo su origen y destino. Cuando un camino cambia para incluir nodos no cooperativos, es notorio el decremento de su reputación.

Tabla IV
CACHÉ PRIMARIA DEL NODO ORIGEN 8 CUANDO TENEMOS MÚLTIPLES CAMINOS.

Posición	Contenido de caché primaria	Camino	Calificación. Sim.
0	[(8) 9 3]	C ₂	0,8724
1	[(8) 9 13]	C ₅	0,8789
2	[(8) 12 13]	C _{5'}	1,0000
3	[(8) 7 6 5]	C ₈	0,7439
4	[(8) 7 6 16]	C ₇	0,7463
5	[(8) 12 11 15]	C ₆	1,0000
6	[(8) 7 1 0]	C ₁	0,7476
7	[(8) 9 10 4]	C ₃	0,8124
8	[(8) 9 10 14]	C ₄	0,8156

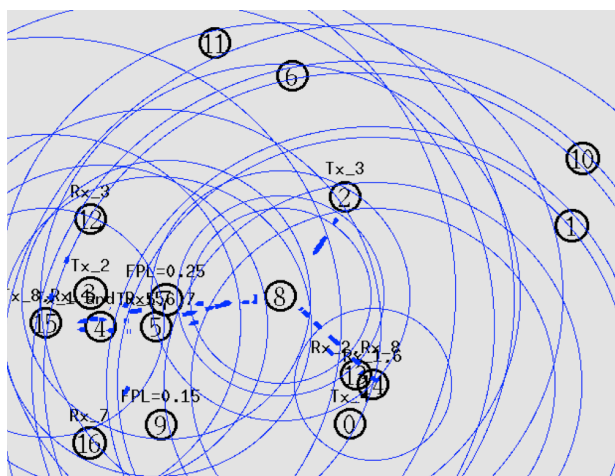


Fig. 8. Escenario con patrón de movilidad *Random Waypoint* a 2 m/s.

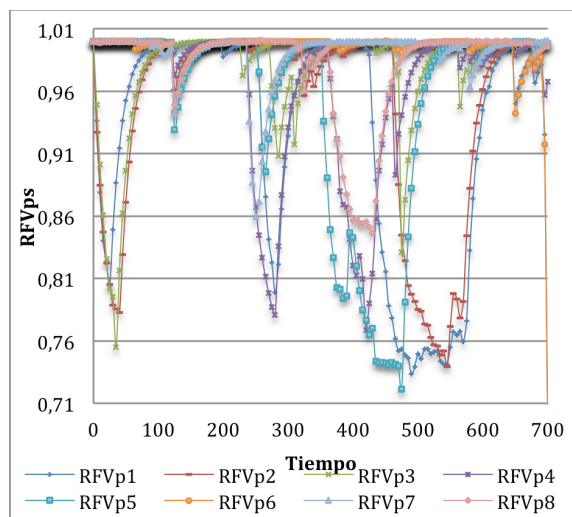


Fig. 9. Reputación de los caminos activos.

La Tabla V muestra los valores simulados promedio de \overline{RFVn}_{j0} (reputaciones de los nodos vistas por el nodo origen 0). Este nodo se comunica con un único destino, el nodo 15, formando el camino 4 (Fig. 2). En la tabla se observa que la mayoría de los nodos han participado en dicho camino de manera que su reputación ha cambiado (disminuido para los nodos colaborativos o aumentado para los egoístas). A pesar de ello los nodos con menor reputación son los que presentan un comportamiento egoísta, siendo el nodo 7 el de peor reputación.

Tabla V
REPUTACIÓN OBSERVADA POR EL NODO ORIGEN 0 CUANDO LOS NODOS ESTÁN EN MOVIMIENTO.

Nodo	Val. Real.	Val. Sim.	Int. Conf. 95% de Val. Sim.
N_0	1	1	(1,0000; 1,0000)
N_1	1	0,9712	(0,9754; 0,9939)
N_2	1	0,9990	(0,9976; 0,9992)
N_3	1	0,9729	(0,9710; 0,9762)
N_4	1	0,9993	(0,9965; 1,0007)
N_5	1	0,9924	(0,9903; 0,9929)
N_6	1	0,9802	(0,9252; 0,9566)
N_7	0,75	0,8837	(0,8899; 0,8986)
N_8	1	0,9993	(0,9931; 1,0002)
N_9	0,85	0,9610	(0,9448; 0,9549)
N_{10}	1	1	(0,9982; 1,0007)
N_{11}	1	0,9681	(0,9501; 0,9620)
N_{12}	1	0,9996	(0,9996; 1,0000)
N_{13}	1	0,9939	(0,9779; 0,9934)
N_{14}	1	0,9791	(0,9721; 0,9840)
N_{15}	1	1	(1,0000; 1,0000)
N_{16}	1	0,9995	(0,9945; 1,0007)

La Fig. 9 muestra la calificación que va obteniendo el camino 4 a lo largo de la simulación. Dicha calificación se inicializa a 1, presentando los picos más bajos cuando participa uno o dos nodos egoístas en el camino. La estabilización a un valor constante diferente de 1 se da cuando se produce una ruptura del enlace entre el nodo 0 y el 15.

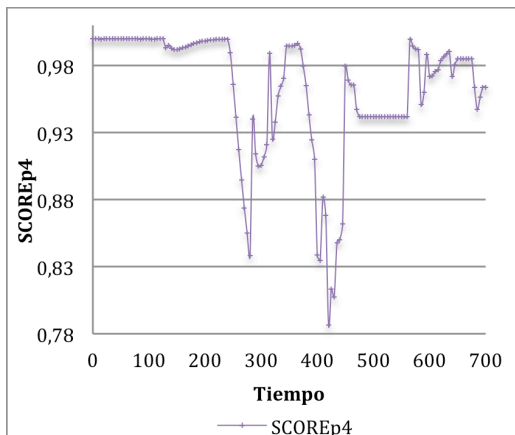


Fig. 9. SCORE del camino activo observado por el nodo 0.

La Tabla VI muestra los valores simulados promedio de $RFVn_{j5}$ (reputaciones de los nodos vistas por el nodo origen 5). En este caso la reputación del nodo 9 se mantiene en 1, debido a que el nodo 9 no participa aún como nodo intermedio en ninguna de las comunicaciones en las que el nodo 5 es el transmisor (caminos 5, 6 y 7). Como el nodo origen 5 es el transmisor en tres caminos distintos, entonces tendrá las calificaciones de las tres rutas activas como lo muestra la Fig. 10.

De esta manera cada nodo origen tendrá las reputaciones de los nodos así como las calificaciones de los caminos, por tanto al igual que se han presentado los datos para el nodo transmisor 0 y 5, los nodos 4, 3, 2 y 15 también calcularán estos valores.

V. CONCLUSIONES

En este artículo se ha propuesto un algoritmo de encaminamiento que permite encontrar las rutas que incluyen a los nodos más cooperativos de una MANET. Dicho

algoritmo utiliza exclusivamente información de primera mano a partir de la experiencia de cada nodo en la red, lo cual simplifica su uso. Cada nodo de la red estima la reputación de reenvío del resto de nodos. Para ello es necesario el conocimiento de las pérdidas en cada camino que utiliza el nodo que hace la estimación, repartiéndolas por igual entre los nodos que forman cada camino. Ello conlleva que en un mismo camino, los nodos que tienen un comportamiento más colaborativo ven penalizada su reputación mientras que los que tienen un comportamiento más egoísta la ven recompensada. Para paliar este efecto, el cálculo se realiza promediando, mediante una media móvil, la reputación estimada, para cada uno de los caminos conocidos en los que el nodo participa. Así pues, la reputación de reenvío de un nodo, no sólo depende de su comportamiento con los reenvíos, sino que además está afectada por el comportamiento de sus nodos vecinos. Este efecto, que a priori puede parecer injusto, facilita el encaminamiento por caminos que incluyan a nodos más colaborativos.

Tabla VI
REPUTACIÓN OBSERVADA POR EL NODO ORIGEN 5 CUANDO LOS NODOS ESTÁN EN MOVIMIENTO.

Nodo	Val. Real.	Val. Sim.	Int. Conf. 95% de Val. Sim.
N_5	1	0,9867	(0,9820; 0,9883)
N_1	1	0,9541	(0,9516; 0,9759)
N_2	1	0,9987	(0,9950; 0,9978)
N_3	1	0,9998	(0,9973; 0,9998)
N_4	1	0,9979	(0,9942; 0,9980)
N_5	1	1	(1,0000; 1,0000)
N_6	1	0,9999	(0,9736; 0,9922)
N_7	0,75	0,9131	(0,9195; 0,9311)
N_8	1	0,9960	(0,9948; 0,9973)
N_9	0,85	1	(1,0000; 1,0000)
N_{10}	1	1	(0,9937; 1,0017)
N_{11}	1	1	(1,0000; 1,0000)
N_{12}	1	0,9998	(0,9909; 1,0005)
N_{13}	1	0,9742	(0,9793; 0,9876)
N_{14}	1	0,9989	(0,9770; 0,9921)
N_{15}	1	0,9996	(0,9972; 0,9999)
N_{16}	1	0,9998	(0,9969; 0,9985)

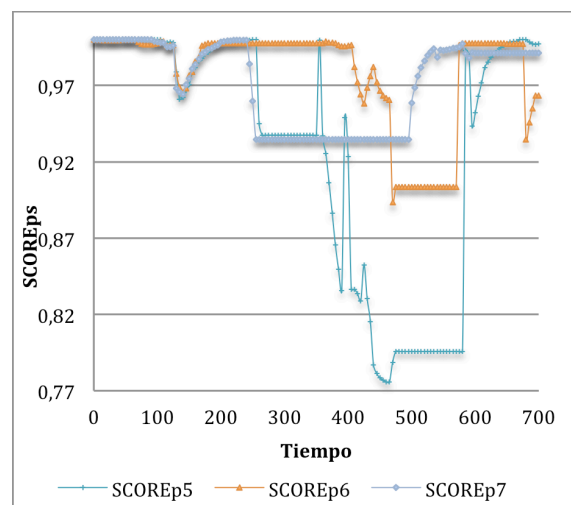


Fig. 10. SCORE para los caminos activos observados por el nodo 5.

La principal desventaja del algoritmo está en el hecho que un nodo con baja reputación participará en menos caminos que uno de mayor reputación, lo cual dificulta que su reputación mejore en caso que el nodo mejore su

comportamiento en la red. Es por ello, que inicialmente la reputación de todos los nodos es la unidad. Una posible solución a este hecho es que los nodos olviden periódicamente las estimaciones de reputación que han hecho y las reinicialicen.

En este trabajo se ha presentado una expresión analítica para calcular la estimación de reputación que realizará cada nodo. Dicha estimación depende de la topología de la red y de los tráficos cursados por los distintos caminos.

Además se ha implementado el algoritmo en el simulador NS-2 para comparar los resultados obtenidos con los resultados teóricos. Se ha modificado el comportamiento de algunos de los nodos de la red para que no reenvíen todos los paquetes que reciben.

En el entorno de simulación se ha analizado la influencia de los parámetros α y β en las medias móviles con ponderación exponencial (EWMA). Valores pequeños de estos parámetros se traducen en una adaptación lenta en el caso que variaciones significativas en el comportamiento de los nodos. Ello también afecta al transitorio inicial de manera que éste es mayor y se tarda más en tener un valor fiable de la reputación de los nodos. Dicho transitorio no parece verse afectado por el valor asintótico al que se tiende, sino únicamente a los valores de α y β . Debido al hecho que la estimación de la reputación de los nodos se hace a partir de la reputación de los caminos, si dicha reputación se actualiza lentamente (valor de α pequeño), la reputación de los nodos también se actualiza lentamente independientemente del valor de β . Por otro lado, valores mayores de estos parámetros dan como resultado valores menos estables en la estimación de la reputación presentando ésta una mayor varianza.

También se han simulado escenarios con múltiples caminos a un mismo destino, los cuales son descubiertos por el protocolo DSR y guardados en la memoria caché del nodo origen. Se ha estimado la reputación de dichos caminos observando que aunque los valores estimados de reputación para cada camino no se corresponden con la reputación real, los caminos con mayor reputación son los que incluyen el menor número de nodos con comportamiento egoísta. Es decir, a pesar que los valores de reputación no son exactos, el camino elegido sí es el que presenta mayor reputación real, lo cual valida el planteamiento del algoritmo propuesto.

Finalmente se ha simulado el modelo en un escenario con movilidad, en donde se ha verificado que los nodos egoístas presentan una reputación más baja que el resto.

AGRADECIMIENTOS

Este trabajo ha sido financiado por los proyectos de investigación TIN2014-58259-JIN (Microagregación Anónima en Encuestas Demográficas a Gran Escala), TEC2013-47665-C4-1-R (EMERGENCY RESPONSE IN SMART COMMUNITIES, EMRISCO), TEC2014-54435-C4-1-R (INCIDENT MONITORING IN SMART COMMUNITIES. QoS AND PRIVACY, INRISCO) y por la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) de la República del Ecuador.

REFERENCIAS

- [1] A. Srinivasan, J. Wu, J. Teitelbaum, "Distributed Reputation-based Secure Localization in Sensor Networks," *Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton*.
- [2] P. Michiardi, R. Molva, "Simulation-based Analysis of Security

- Exposures in Mobile Ad Hoc Networks," *Institut Eurécom, France*.
- [3] J. Muñoz, O. Esparza, M. Aguilar, V. Carrascal, J. Forné., "RDSR-V. Reliable Dynamic Source Routing for video-streaming over mobile ad hoc networks.," *Computer Networks*, no. 54, p. 17, 2010.
- [4] P. Chandra, D. Dobkin, A. Bensky, R. Olexa, D. Lide, F. Dowla, *Wireless Networking*. Oxford: Elsevier, 2008.
- [5] D. Johnson, D. Maltz. (2007, February) RFC:4728. [Online]. www.ietf.org/rfc/rfc4728.txt
- [6] A. Srinivasan, J. Teitelbaum, J. Wu, M. Cardei, H. Liang, "Reputation-and-Trust-Based Systems for Ad Hoc Networks," in *Algorithms and Protocols for Wireless and Mobile Ad Hoc Networks*. Hoboken, New Jersey: John Wiley & Sons, 2009, pp. 375-403.
- [7] S. Bansal, M. Baker, "Observation-based Cooperation Enforcement in ad hoc Networks," *Stanford, CA, USA, 2003*.
- [8] S. Buchegger, J. Le Boudec, "Performance Analysis of the CONFIDANT protocol (Cooperation Of Nodes: Fairness In Dynamic Ad-hoc NeTworks)," *MOBIHOC'02*, June 2002.
- [9] P. Michiardi, R. Molva, "CORE: A Collaborative Reputation Mechanism to enforce node cooperation in Mobile Ad hoc Networks," *Sophia Antipolis, France*.
- [10] K. Fall, K. Varadhan, *The ns Manual (formerly ns Notes and Documentation), The VINT Project*, 2011.
- [11] (2013, July) BonnMotion. A Mobility Scenario Generation and Analysis Tool. [Online]. http://sys.cs.uos.de/bonnmotion/doc/BonnMotion_Docu.pdf

Multi constraint network virtualization strategy based on Paths Algebra for non-linear parameters

Xavier Hesselbach^a, José-Roberto Amazonas^b, Alejandro Gutiérrez-Anato^a

^a Universitat Politècnica de Catalunya UPC, C/Jordi Girona,31, Barcelona, Spain

^b University of São Paulo USP, São Paulo, Brazil.

xavier.hesselbach@upc.edu, jra@lcs.poli.usp.br, anatoguti@gmail.com

Abstract- One of the main challenges in network virtualization is the virtual network embedding problem (VNE). The objective of the VNE is to map a set of Virtual Network Requests (VNR) to physical nodes and links. VNR is composed by a set of virtual nodes and links with several demands (traditionally CPU and bandwidth) to be assigned into a set of paths in the substrate network with sufficient resources. Embedding can be optimized with regard to specific metrics. To help solving the VNE problem a mathematical tool has been proposed, called Paths Algebra, to deal with the VNE multi-constraint problem using linear metrics. Most of the existing VNE proposals treat the single-path virtual link-mapping problem as a mono-constraint. This paper studies, analyses and validates the mathematical framework of Path Algebra under selected linear (CPU, bandwidth and energy consumption) and non-linear metrics (availability and packet loss ratio), providing promising performance results.

Keywords- Network Virtualization, Paths-Algebra, Virtual Network Embedding, Mapping, Linear and Nonlinear metrics.

I. INTRODUCTION AND OBJECTIVES

Network virtualization will be a fundamental technology in IaaS (Infrastructure as a Service) by allocating several virtual networks (VN) on top of a substrate network (SN). The effective allocation of these resources is the objective of the VNE (Virtual Network Embedding) problem.

Most of the existing VNE proposals treat the single-path virtual link-mapping problem as a mono-constraint, that is, their objective is to map the virtual links in substrate paths that minimize/maximize the usage of one resource (typically bandwidth). This paper introduces a virtual link mapping approach supporting multiple constraints using the Paths Algebra framework with non-linear parameters.

VNE link mapping corresponds to finding the best route(s) on the substrate network for each virtual link in the VN, under a defined optimization criteria.

In previous work [1], it was already proposed the utilization of the Paths Algebra to solve the virtual link mapping stage of the VNE. Paths Algebra provides the flexibility to introduce an unlimited number of linear and non-linear constraints and metrics to the problem while finding all the eligible paths in the physical network resulting in better and more flexible embedding.

For now, this mathematical tool has been used to solve VNE with linear metrics, such as CPU, number of hops or bandwidth (BW). The main objectives in this work is to study, analyze and validate the mathematical framework of Path Algebra, how the mapping is achieved and how the physical constraints and selected metrics affect the mapping of the Virtual Network Requests (VNRs) with non-linear and linear

metrics. This is analyzed in the simulation environment in different scenarios.

II. BACKGROUND

A. Virtual network embedding

As it was stated before there is a challenge in trying to allocate the different resources of the substrate network to the VNRs: the VNE problem. Through dynamic mapping of virtual resources onto a physical hardware, the benefit gained from existing hardware can be maximized. Optimal dynamic resource allocation, leading to the self-configuration and organization of future networks, will be necessary to provide customized end-to-end guaranteed services to end users [2].

This optimal allocation can be done to reach certain objectives, for instance Quality of Service or economical profit, energy-efficiency, etc.

The virtual network embedding can be solved in two stages: the first one is the Virtual Node Mapping [2] which is responsible of allocating virtual nodes of the VNR in the physical nodes with enough capacity to accomplish the virtual node resource demand, and the Virtual Link Mapping [2] where the virtual link connecting these virtual nodes are mapped to a directed path in the substrate network with enough resource capacity to meet the virtual link demand.

Two main approaches exist in the context of VN embedding depending on the considered scenario: in the offline or static approach, all the VN demands are known in advance and the objective is to maximize the number of correctly mapped demands over the physical substrate given the scarcity of resources. In the online or dynamic approach, demands are not known in advance but they arrive (or leave) dynamically at the network [3].

Formally the VNE can be described as follows: a SN is represented by a direct graph, with nodes (N) and a links (L), $S=(N,L)$, then on top of the substrate network a set of virtual network request (VNR), $VNR^k=(VN^k,VL^k)$, where k are the set of virtual request $k=1,2,3\dots$

Two functions are then defined: a node mapping and a link mapping. For the node mapping is defined $X:VN^k \rightarrow N$ to assign virtual nodes to substrate nodes. Likewise the function $Y:VL^k \rightarrow 2^P$ where 2^P consists of all sets of directed paths in the SN. If Y is able to assign a virtual link to a set with more than one element, the VNE problem will allow a virtual link mapping with the use of multi-path routing [3].

For both the node mapping and link mapping function the resources of the substrate node are limited, so is important that neither exceed the resources of either a node or a link in the

SN. An optimal VNE is then the result of the node and link mapping functions that satisfies all of the restrictions and, additionally, reaches a given optimization objective.

B. VNE state of the art

Some solutions have been already proposed in the past. The most relevant contributions concerning this work are the following: two algorithm were proposed by Zhu and Ammar in [4], the first one called “Basic VN assignment algorithm” to do a minimization of the node and link stress sum, and the second algorithm called “Subdividing Algorithm”, splitting each VNR into a set of connected sub-VNs, each with a star topology.

The second approached was proposed by J. Lu, J. Turner [5] in which they only consider the bandwidth constraint and is evaluated in an offline scenario. In this approach, VN topologies are limited to be “backbone-star” topologies. These topologies are composed of “access nodes”: nodes representing the SN location at which traffic enters to the VN, and “backbone nodes” connecting the access nodes. An iterative algorithm changing the mapping of backbone nodes is performed to embed the different VNRs.

Other approach by Yu et al. in [6] looks for the maximization of long average revenue i.e. the weighted sum of VNR’s bandwidth and CPU demands. CPU in nodes and BW in links are the resources considered in this proposal

Chowdhury and Boutaba propose an approach to coordinated node and link mapping in [7]. Their main goal is to minimize VN embedding cost. VNE is performed in just one stage in Lischka and Karl [8] proposal. Node and link mapping are performed in the same stage by reducing the VNE to the well-known NP-complete sub-graph Isomorphism Detection (SID) problem.

At this point, the concept of Hidden Hops permits more flexible selection of allowed paths. The concept was first introduced in [9]: a substrate node is working as a hidden hop if it is part of a SN path used to map a virtual link (i.e. it is an intermediate node in the SN path). Besides the demands of some virtual nodes, each hidden hop will have a demand of resources because it has to be configured to forward packets passing through the SN path.

As it was pointed out before, a procedure to get a solution for the virtual mapping problem is to divide in two stages: node mapping and link mapping. An optimal exact solution was proposed by Houidi et al. [10]. The solution is found by solving a MIP optimal problem formulation. Although optimal embedding can be performed in this way, as MIPs are complex problems (NP-complete), this solution is not scalable and it will take unaffordable running times for large networks. However, it can be used as a baseline for new VNE heuristics.

All these previous proposals address the VNE problem in a single Infrastructure Provider (InP) scenario (so called Intra-domain VNE).

However, there are proposals to perform the VNE in an inter-provider scenario. For instance Chowdhury et al. in [11] propose a policy-based end-to-end VNE framework (PolyViNE) to embed VNs across multiple InPs. Inside any InP, any of the previous intra-domain proposals can be used, but across domains a distributed protocol that coordinated the participating InPs ensuring competitive pricing is required.

C. Paths Algebra framework

To accomplish a Virtual Network Mapping with a QoS control and QoS parameters (delay, losses, availability, distance, flow, etc.), several network routing solutions can be proposed, such as exact and approximated routing algorithms, algorithms based on backward-forward heuristics, linear composition, hybrid, random, routing computation from the origin or destination, reservation and resource allocation protocols [12,13,14,15].

In these solutions, their validity is often defined either by the size of the networks or by their topology.

Analyzing this problem under the perspective of protocols design, it has been necessary to conceive an heuristic or an algorithm to ensure the routing convergence for different types of QoS metrics or QoS metrics composition, in which this problem could be addressed from an integrated and generic manner by means of a mathematical framework, validating the proposed solutions independently from network topology or implementation details [16].

A framework considered in which the user can define their own path searching policy that can be closer to the existing traffic profile of their networks is Paths Algebra [17]. Such mathematical framework allows for systematically comparing different mono-constraint and multi-constraint routing heuristics concerning their convergence guarantees, best path convergence and loop avoidance, and validating a generic and homogeneous solution that can be integrated into a single mathematical. In this case, QoS is a function of physical parameters associated to the network’s performance [18].

Considering Network Virtualization mainly as a business strategy, the main metrics to be use would be based in cost and revenue. Nevertheless, Paths Algebra may also use these metrics and, for example, associate them with the QoS. Such flexibility associated to its computing efficiency makes the Paths Algebra a suitable tool to perform the virtual link mapping stage inside the VNE [1].

Using Paths Algebra, the virtual link mapping can be performed in a multi-constraint basis, that is, virtual links mapped to substrate paths, characterized by an unlimited number of constraints (or combination of constraints). In previous work [1] this framework was tested using two metrics: the CPU of each node and the bandwidth of the links. One of the contributions in this paper is the inclusion of non-linear metrics and the behavior analysis using cost, revenue, the mapped revenue ratio and the acceptance ratio.

D. Paths Algebra overview

Paths Algebra is defined and executed in 3 main steps: the combination metrics, the synthesis function, and the ordering.

To this end, a network is represented as a graph $G=(V,A)$, where V are the vertices or nodes and A as the set of arc or links. Considering the simple network in the Fig.1, the vertices are given by $V=(1,2,3,4,5)$ and the arcs are $A=(a,b,c,d)$. For instance, the path $p_{a,d}$ from the source node 1 and destination 5, is (a,b,c,d). In other words, this path is represented as a successions of vertices.



Fig. 1. Example of a path p_{ad} in a network.

Each link in the network can be characterized by the following set: $(m_1(x), m_2(x), f[m_1(x), m_2(x)])$, where $m_1(x)$ and $m_2(x)$ are the metric values considered in every link x , and $f[m_1(x), m_2(x)]$ is a combination of both metrics.

For this example the following matrix $C(p_{a,d})$ in (1) represents the combination of all of the metrics to consider in the path of the Fig. 1:

$$C(p_{a,d}) = \begin{pmatrix} C_a \\ C_b \\ C_c \\ C_d \end{pmatrix} = \begin{pmatrix} m_1(a) & m_2(a) & f[m_1(a) + m_2(a)] \\ m_1(b) & m_2(b) & f[m_1(b) + m_2(b)] \\ m_1(c) & m_2(c) & f[m_1(c) + m_2(c)] \\ m_1(d) & m_2(d) & f[m_1(d) + m_2(d)] \end{pmatrix} \quad (1)$$

The so called Synthesis $S[]$ is defined in Paths Algebra as a set of operations applied on the values of the links metric combination $C(p_{a,d})$. By using this formula, a characterization of the path based on the constraint is done. The synthesis used could be, for example, $\{add(), mult(), max(), min()\}$.

If the routing has only one constraint, only one value would be obtained. On the other hand, if the routing is multi-constraint with k -constraints, then k values are obtained.

So, the synthesis can be formulated as follows:

$$S[] = [S_1 \ S_2 \ S_3] \quad (2)$$

where the first value S_1 corresponds to resulting value of the synthesis applied to the whole path; the second (S_2) corresponds to resulting value of the synthesis applied to the sub path obtained by dropping out the last arc; until the last one (S_3 in formula (2)) corresponds to the resulting value of the synthesis applied to the sub path constituted only by the first arc.

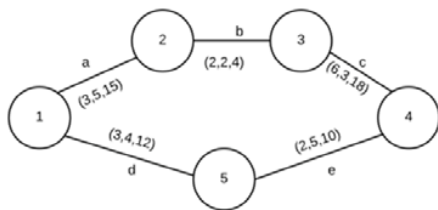


Fig. 2. Path ordering example.

Once this is done, a path ordering is executed in the following way: consider the network represented in Fig. 2, there are two different paths between the nodes 1 and 4. The first path is the one with the nodes (1,2,3,4) and the links (a,b,c). The second path is with the nodes (1,5,4) and the links (d,e). Each link in the path is characterized by the triple formulation explained before $\{m_1(x), m_2(x), f[m_1(x), m_2(x)]\}$, where, for example, the following is considered:

$$f[m_1(x), m_2(x)] = m_1(x) * m_2(x) \quad (3)$$

with the next vector synthesis:

$$S[] = [\min() \ max() \ add()]^t \quad (4)$$

Once the synthesis is calculated, the ordering of the paths is ready to be performed using a multidimensional lexical (ML) ordering. Different ordering criteria can be considered, upon the specific problem considered: $ML \leq = \{\geq, \leq, \geq\}$. For instance:

if $S[\text{Path1}] \leq S[\text{Path2}]$ then Path1 will be less optimal that Path2.

III. PROPOSAL

As demonstrated in [1], Paths Algebra can help to solve the virtual link mapping stage of the VNE problem. As explained in the previous section, Paths Algebra provides the flexibility to introduce an unlimited number of linear and non-linear constraints and metrics to the problem while finding all the eligible paths in the physical network resulting in better and more flexible embedding.

In this work, new constraints are analyzed including 3 new non-linear and linear metrics introduced in the path computation optimization problem: Packet Loss Ratio, Availability and Energy Consumption. The main idea is to study and analyzed how Paths Algebra framework can help solving the VNE problem under this new scenario.

The strategy will be simulated with the tool ALEVIN [19] (Algorithms for Embedding Virtual Networks).

In order to define metrics, the network topology shown in the Fig. 3 is used, since it is small enough for manual computation. For simulations, bigger scenarios are considered, as explained in section IV.

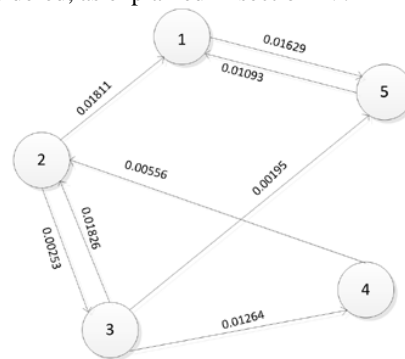


Fig. 3. Substrate network topology, and the Packet Loss Ratio.

Table 1. Values considered for the PLR adjacency matrix

	1	2	3	4	5
1	0	0	0	0	0.01629
2	0.01811	0	0.00253	0	0
3	0	0.01826	0	0.01264	0.00195
4	0	0.00556	0	0	0
5	0.01093	0	0	0	0

A. Metrics

The first metric analyzed is the Packet Loss Ratio (PLR). This metric is non-linear and measures the percentage of packets lost in a network.

The range generated during simulations will be between 0% and 2%, for the links of the substrate network. In order to include the metric into the ALEVIN environment, the next procedure is followed: Using the adjacency matrix M of the substrate network, a new matrix is created for the Packet Loss Ratio. So when the element $a[i,j]$ of the adjacency matrix M is equal to 1, the link with the source node i and destination j will have a Packet Loss Ratio (PLR) value.

Once the PLR is created, a synthesis must be used in order to get a value characterizing the different paths. For the PLR, is used 1-PLR, the synthesis associated is multiplication $mult()$ and the ordering relation is \leq .

Let's consider Fig.4 and Fig.5, defining two paths from nodes 2 to the destination node 5 with different PLR values, as taken from Fig.3. Path 1 in Fig.4 is composed of the nodes 2, 1, 5 and Path 2 in Fig.5 is composed of the nodes, 2, 3, and 5.

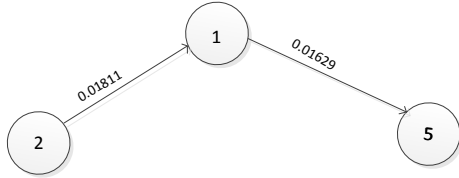


Fig. 4. Path1 Packet Loss Ratio

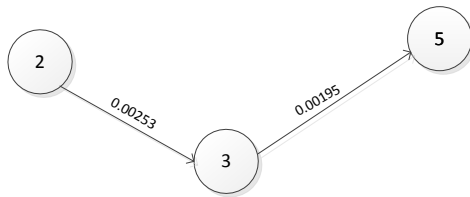


Fig. 5. Path 2 Packet Loss Ratio

Calculating the end-to-end losses probability and using the synthesis, the value for each path is the following:

$$PLR_{Path1} = 0.0341 \text{ (3.41\%)} \\ PLR_{Path2} = 0.0044 \text{ (0.44\%)}$$

The next metric considered in this work is the Availability, defined as the percentage value of the amount of time the network is delivering services divided by the amount of time it is expected to deliver services.

Using the same adjacency matrix M of the substrate network than before, a new matrix is created for the Availability: When the element $a[i,j]$ of the matrix M is equal to 1, the link with the source node i and destination j will have an Availability value. For instance, the values of links in the substrate network of Fig.5, in a matrix called A, where the element $Aval[i,j]$ represents the value of the Availability of each link in the SN, are shown in table 2.

Table 2. Values considered for the Availability adjacency matrix.

	1	2	3	4	5
1	0	0	0	0	0.99787
2	0.99824	0	0.95788	0	0
3	0	0.99852	0	0.99785	0.97426
4	0	0.99001	0	0	0
5	0.95709	0	0	0	0

As in the case of PLR, once the metric is created, a synthesis must be used in order to get a value to characterize the paths. For the Availability, the synthesis associated is multiplication $\text{mult}()$ and the ordering relation will be \leq . To illustrate better this explanation, the following example is presented:

Consider 2 paths between the s node 2 and the destination node 5 with different availability (Fig. 6 and Fig.7). Path 1 is composed of the nodes 2, 1, 5 and the Path 2 is composed of the nodes 2, 3, 5.

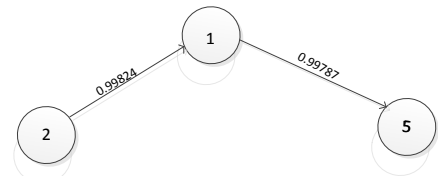


Fig. 6. Path 1 availability

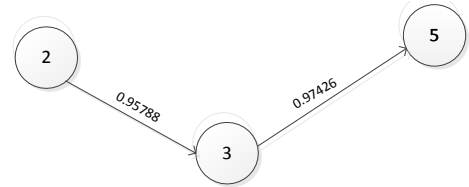


Fig. 7. Path 2 availability

Using the synthesis and ordering relations, the corresponding value for each path are the following:

$$Availability_{Path1} = 0.9332 \text{ (93.32 \%)} \\ Availability_{Path2} = 0.9961 \text{ (99.61 \%)}$$

The last metric included in this paper is the Energy consumed by the nodes in the substrate network.

In order to create the energy metric, the next simple model will be used, regarding models from [20], where a fix consumption is considered (A) plus a linear contribution from the BW parameter:

$$Energy = A + B \cdot BW \tag{5}$$

BW is the maximum inflow and outflow bandwidth of every node in the network and A and B are constants that depends on the energy consumption of the hardware of the nodes. This simple model based on the valued of A and B can balance the energy consumption between the chassis of the nodes and the node processor consumption.

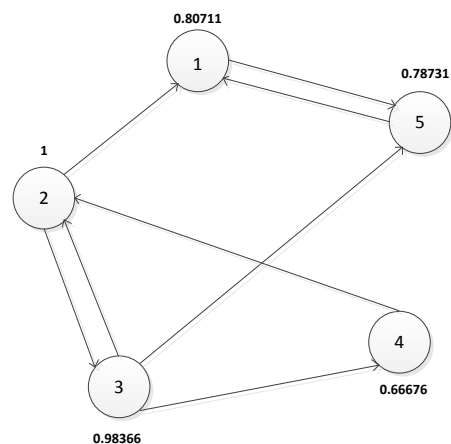


Fig. 8. Energy consumption in the SN.

As in previous cases in order to use this metric, a vector is generated. Each element in the vector will correspond to the value of the energy in each node. For instance, the energy consumption of each node in Fig.8 is presented in the table 3.

Table 3. Values considered for the energy consumption per node (units are not relevant in this example, Wh, KWh, etc.).

1	2	3	4	5
0.8071	1	0.9836	0.6667	0.7873

For the Energy metric the synthesis associated will be adding add () and the ordering relation will be \geq , in order to minimize the consumption. Considering two paths between the origin node 2 and the destination node 5 with different energy consumption, and using the above procedure, the consumption for each node is computed and processed using the synthesis function.

Consider for example the part of a network shown in Fig. 9. In this figure, the weights are associated to the nodes. Energy consumption of nodes A, B and C are given by Energy(A), Energy(B) and Energy(C), respectively.

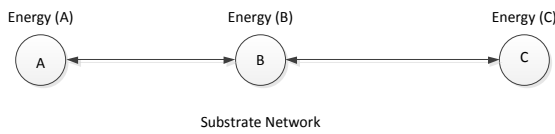


Fig. 9. Example of energy consumption in a SN path.

As the objective is to minimize the energy consumption, the new energy metric can use the metrics combination function provided by the paths algebra and reduce the metrics to (sum [Energy(A), Energy(B)], Energy(C)) as shown in Fig.10.

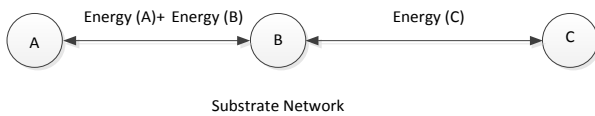


Fig. 10. New metric combination.

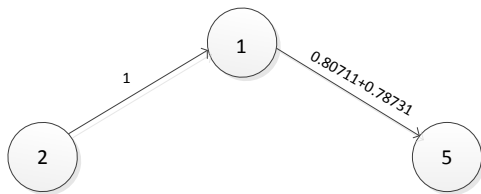


Fig. 11. Path 1 energy consumption

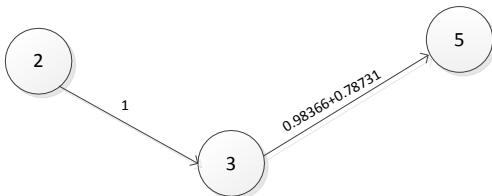


Fig. 12. Path 2 energy consumption.

The value for each path is the following:

Energy Consumption Path1: 2.5944
 Energy Consumption Path2: 2.7709

IV. EVALUATION SCENARIO

The proposal and metrics are evaluated using the ALEVIN software [19], an efficient graphical tool oriented to VNE problems implementing several algorithms and metrics. The goal is to analyze if Paths Algebra is able to map efficiently the different links of the VNR into the SN using the new metrics priority defined in this work.

A. Network generation

For each of the scenarios, the Waxman topology is considered [1]. The ALEVIN software is able to generate the substrate network topology, using the following formula:

$$P(u, v) = \alpha e^{-\left(\frac{d(u,v)}{\beta \cdot L}\right)} \quad (6)$$

with $0 < \alpha, \beta \leq 1$. Larger values of β result in graphs with higher link densities, while small values of α increase the density of short links relative to longer ones. d is the Euclidean distance, between the nodes v and u , and L is the maximum Euclidean distance between any pair of nodes. An increase in α will increase the probability of edges between any nodes in the graph. This topology generation procedure is used for the substrate network as well as for all virtual networks.

For the scenarios considered in this work, the values of α and β are 0.5. These values are chosen in order to provide a network with relative medium links density. Furthermore, as the coordinates of the nodes are going to be uniformly distributed in a 1x1 square area, the average distance between two nodes will be 0.5 and the maximum distance $\sqrt{2}$. Thus according to the Waxman equation (6), the average probability creating a link between any two nodes is $\frac{1}{4}$. The load for resources and demands will also be generated randomly. The exact generation method is specified in [1].

B. Scenarios

4 scenarios are compared using the new nonlinear and linear metrics, as included in the ALEVIN environment. Each scenario has different metric priority (Table 4).

Table 4. Scenarios evaluated.

Scenario	Optimization Criteria	Metric Ordering	Metric Priority	Physical Constraint
1	Minimize Cost	Hops, Bandwidth, CPU	Hops	BW, CPU
2	Maximize Throuput	PLR, Hops, Bandwidth, CPU	PLR	BW, CPU
3	Maximize Availability in the Substrate network	Availability, Hops, Bandwidth, CPU	Availability	BW, CPU
4	Minimize Energy consumption	Energy, Availability, Hops, Bandwidth, CPU	Energy	BW, CPU

C. Simulation parameters

Table 5 presents the parameters for each scenario. The values chosen provide effective runtime of the algorithm and a realistic scenario. The load range from 0.1 (lightly loaded) through 0.99 (heavily loaded). N = 20 runs are performed for each value of load in order to include randomness in the simulation.

Table 5. Simulation parameters

Parameters Descriptions	Value
Nodes in the substrate network	20
Nodes in the Virtual Networks	10
Number of Virtual Networks (k)	10
Loads (ρ)	0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.95,0.99
Number of Runs per Load (N)	20

The total number of times that each algorithm has been performed is given by:

$$N \cdot k \cdot (\text{number of loads analyzed}) = 20 \cdot 10 \cdot 11 = 2,200$$

Also, since this work is comparing 4 scenarios, the total number of VNR assignment attempts is 8,800. Consequently, the execution is made in a high performance computer. The simulation size will be large enough to provide quality comparison between the scenarios and to help understand the behaviors.

D. Simulation metrics

From relevant works of the literature [1, 2, 7, 10, 20], the following metrics are considered in the simulations:

- Average value of Packet Loss Ratio is the Packet Loss Ratio value of each path mapped divided by the total number of the paths being mapped.
- Average Value of Availability: the Availability value of each path mapped divided by the total number of the paths being mapped.
- Average Value of Energy: the energy consumption of each path mapped divided by the total number of the paths being mapped.
- Average Value of hops: the value of the hops of each path mapped divided by the total number of the mapped being mapped.
- Cost/Revenue(C/R): the lower this relationship is, the better the result. Minimum is 1, meaning that is consumed exactly what is required.
- Percentage of Accepted VNR (Acceptance Ratio): the percentage of virtual networks successfully mapped.
- Mapped Revenue Ratio: the percentage of mapped revenue over the total revenue that could be mapped.

V. RESULTS

From the simulations, the following figures were obtained for the four scenarios in each graph (“hops”, “PLR” “availability” and “energy”, respectively):

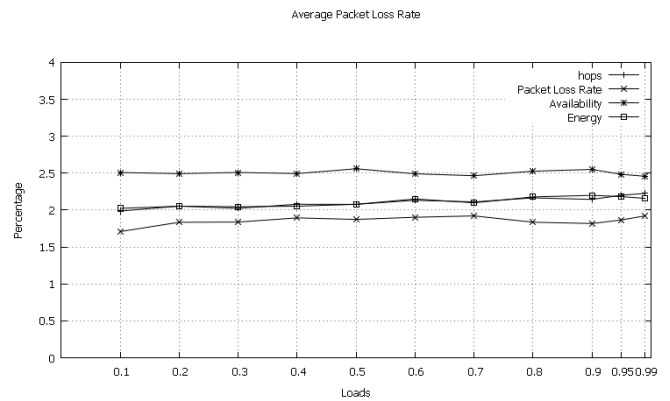


Fig. 13. Average PLR

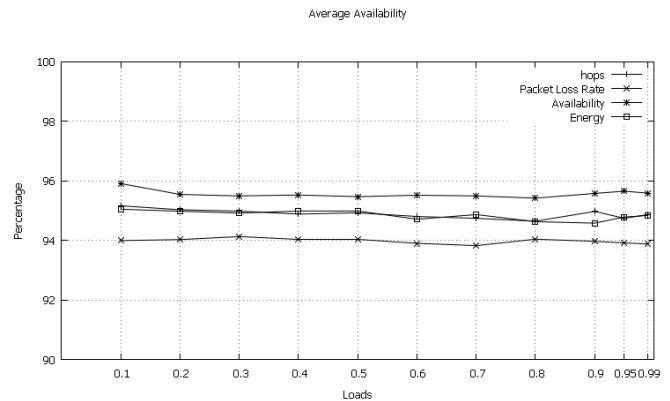


Fig. 14. Average Availability

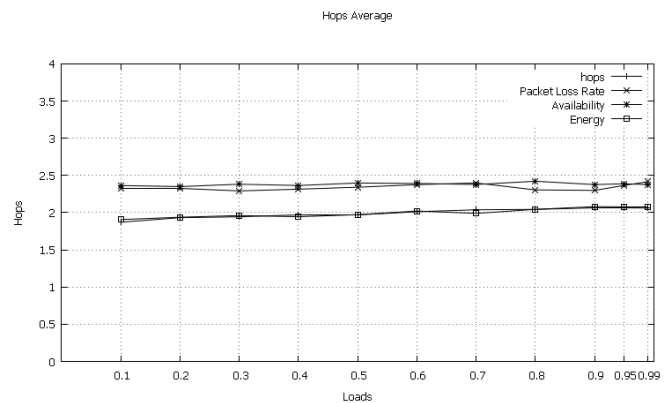


Fig. 15. Average Hops

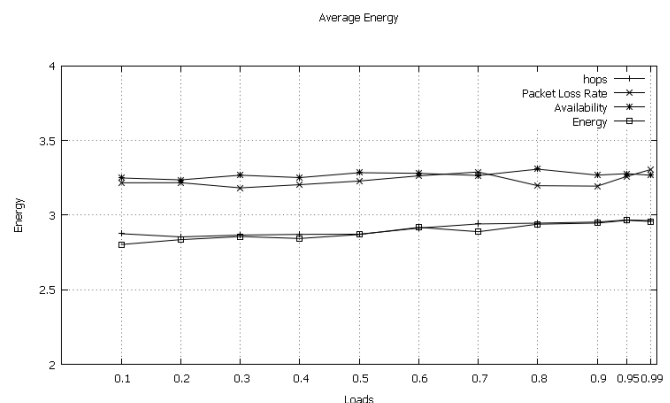


Fig. 16. Average Energy Consumption

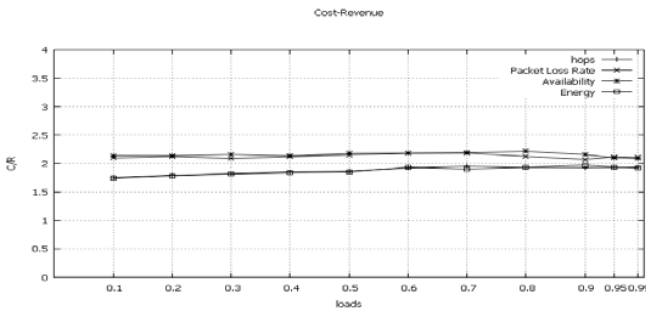


Fig. 17. Cost/revenue

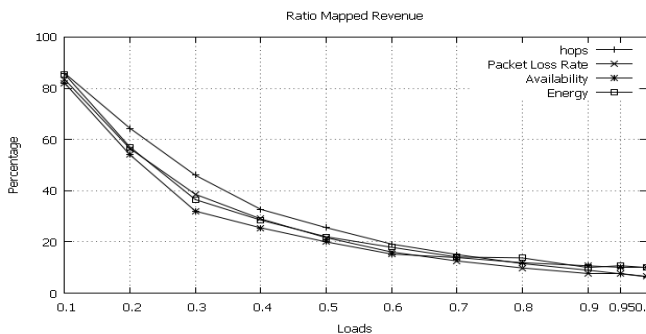


Fig. 18. Mapped Revenue Ratio

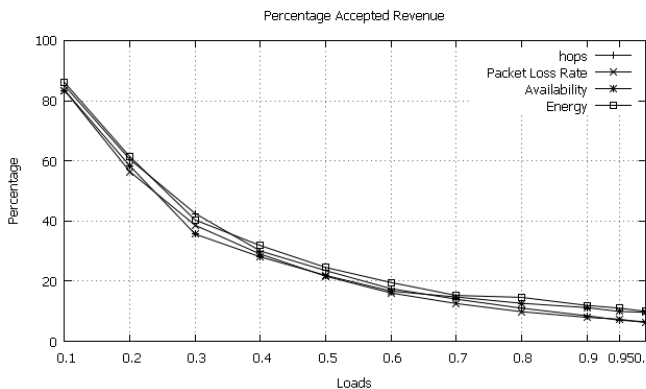


Fig. 19. Percentage Accepted VNR

The observed behavior in Fig.13 comes from the priority of the PLR in scenario 2, where Paths Algebra try to choose the paths of the SN with the lowest Packet Loss Ratio. So, calculating the PLR average of the paths mapped for scenario 1, the value will be the lowest. So it can be concluded that Paths Algebra is choosing the best option. For scenario 2 in which the priority is the Availability, Paths Algebra do not consider the value of the PLR, so the paths were chosen based on the availability of the links. In consequences the PLR is the highest. Finally, for scenario1 and 4, PLR is also not considered, but as the priority are hops and energy, the amount of hops of the paths mapped are shorter that in the case of scenario 3, so the value of PLR is still small.

Concerning Fig.14, in the scenario 3 the average of the availability is the highest. On the other hand for the other three scenarios (Hops, PLR and energy) the value is lower. When Availability is the priority (scenario 1), Paths Algebra has mapped the paths with highest availability, hence, the average

value of the availability for this scenario will be the highest, as shown in Fig.14. On the contrary for the other scenarios the ordering is different, since availability has not been considered as the metric with the biggest priority, so the chosen paths are based only on the priority of each scenario, and consequently the availability is lower.

Fig.15 shows that the averages of the hops are the lowest for the first scenario. On the other hand for the second and third scenario the average of hops are the worst. For the fourth scenario (“energy”) the value of hops are very similar to the scenario 1, as is expected to be.

Regarding energy consumption, Fig.16, results are very similar for scenario 1 and 4, because while trying to reduce the energy consumption of the paths, the amount of hops is also optimized. At the other hand, for the scenario 2 (Packet Loss Rate) and scenario 3 (availability) the value of the energy consumption is the highest. This is because in both scenarios, energy is not the priority and Paths Algebra maps the path with no energy based priority. Consequently the energy consumption is higher.

From the Cost/revenue Fig.17, given that the Paths Algebra try to discover the paths with less amount of energy consumed, these paths will be the shorter. When the load increases, in order to satisfy the VNRs, longer paths have to be used and the cost will increase, so for the case of energy metric and hops an increase of the C/R will be produced. However, the acceptance ratio curve must necessarily be considered in order to understand the behavior, since this value decreases with the load of each network, affecting the cost, the revenue, the cost/revenue and the mapped revenue ratio. On the other hand when availability and PLR has the highest priority, the cost is bigger. Therefore, the paths chosen are larger. This also can be seen from figure 15: that graph shows that for scenarios 2 and 3 the average of hops is larger. That is because in the case when both metrics has the priority, Paths Algebra do not try to find the shorter path, but the paths that have the less packet loss rate or the bigger availability. Another conclusion to be extracted from this figure is the behavior for loads higher than 0.9, since from this point the value of the C/R for the four scenarios converge. This is because for extremely big loads, few resources are available and the strategies will try to choose whichever paths it can, not matter what the metric priority is.

It is worth to conclude from Fig. 19 that the percentage of accepted VNR is the same in all the scenarios when the loads are small. And as the load increases the value decreases due to the lack of resources. When the load is small, enough resources are available to satisfy the demand and all VNRs can be successfully mapped. As the load per Virtual Network increases, the VNR’s are demanding more resources, the load cannot be mapped and the VNR will be dropped.

VI. CONCLUSIONS AND FUTURE WORK

This paper has analyzed the challenges of network virtualization to provide efficient allocation of virtual network elements on top of SN elements, regarding the complexity of the VNE problem under a combination of non-linear and linear parameters. In previous works in the literature, only linear

parameters were considered in practical scenarios. The Paths Algebra has proved its ability to provide results efficiently in an affordable running time using multi-constraints with non-linear parameters. In this work, three new metrics have been included: Packet Loss Ratio, availability and energy consumption. Other works in the literature, so far only has focused in 3 physical constraints: the hops in the SN, the bandwidth and the CPU of each node. Even considering this, since the ALEVIN environment used in this work requires enormous amount of computational resources and regarding the increasing complexities for big network, the networks size and the length of the paths to be analyzed were limited.

It is demonstrated that the optimized metrics obtained in this work are directly connected with the scenario priority considered, as intuitively expected. So for instance for the average of PLR, the smallest value obtained is for the scenario having the PLR as the metric with the highest priority. Same situation for availability, since the value is the highest for scenario 3 and also it is the highest for the hops and energy scenarios. For the case of the C/R the best value occurs for the "hops" and "energy" scenarios. This happens because the metric priority chosen to map the VNR is the shortest available path. For shorter paths, the cost will be smaller and consequently the C/R will also be smaller. On the other hand the worst values are for the second "PLR" and third scenario "availability". This is justified because the paths chosen are larger. Lastly for the graph regarding the Percentage of Accepted VNR and the ratio mapped revenue, the values are similar for the all the scenarios when the loads are small, but as the value decreases the amount of accepted VNR also decreases. This is because if we increase the demands of resources of the VNR, the SN won't have enough resources to accommodate all the VNR, so, some will be drooped and the percentage of accepted VNR will be smaller. Finally these results show that each mapping decision was based on the metric that were prioritized, so the allocation of resources were based on the ordering of the metric set by the users. It can be concluded that Paths Algebra is validated as a powerful and flexible tool for linear and non-linear parameters, in order to optimize and also to define heuristics.

Concerning future works, a backtracking strategy is recommended. The idea behind this method is the following: if a bad mapping decision is detected, a back-track to the previous valid mapping decision is made, avoiding a costly re-map. This will optimize the amount of Virtual Network attended and also will help in allocating the resources in an effective way. Another enhancement will come from the inclusion of an energy model in which the path selection can be done based on the equipment or nodes that are under-utilized. With this solution the energy consumption could be reduced and the model will be more closed to the reality. Also, simulations using a different strategy to increase the load will be very interesting: In this work, the number of VNR remains static, and the load of each increases. In future works, the load of each VN can be static and the number of VNs can be increased with the load. This analysis will impact the cost, revenue and acceptance ratio. At least but not less important, is also proposed the creation of a QoS function to include it in the Paths Algebra based on different linear or linear metrics, in order to take QoS aware embedding decisions.

ACKNOWLEDGEMENTS

This work is partially supported by the project funded by the "Ministerio de Economía y Competitividad" of the Spanish Government, "Redes troncales y de acceso inteligentes definidas por software" TEC2013-47960-C4-1-P.

REFERENCES

- [1] J. F. Botero, X. Hesselbach, M. Molina, J.R. Amazonas, "A novel paths algebra-based strategy to flexibly solve the link mapping stage of VNE problems", *Journal of Network and Computer Applications*, Volume 36, Issue 6, November 2013, pp. 1735–1752.
- [2] Fischer, A; Botero, J.F.; Till Beck, M.; de Meer, H.; Hesselbach, X., "Virtual Network Embedding: A Survey," *Communications Surveys & Tutorials*, IEEE, vol.15, no.4, pp.1888,1906, 2013. DOI 10.1109/SURV.2013.013013.00155
- [3] A; Perelló, J.; Spadaro, S., "Virtual network embedding in optical infrastructures," *Transparent Optical Networks (ICTON)*, 2012 14th International Conference on, vol., no., pp.1,4, 2-5 July 2012. DOI 10.1109/ICTON.2012.6253704
- [4] Y. Zhu, M. Ammar, "Algorithms for assigning substrate network resources to virtual network components", in: *Proc. IEEE INFOCOM*, 2006, pp. 2812–2823.
- [5] J. Lu, J. Turner, "Efficient mapping of virtual networks onto a shared substrate", Washington University in St. Louis, Tech. Rep., 2006.
- [6] M. Yu, Y. Yi, J. Rexford, M. Chiang, "Rethinking virtual network embedding: Substrate support for path splitting and migration", *ACM SIGCOMM CCR* 38, 2008, pp.17–29.
- [7] N. Chowdhury, M. Rahman, R. Boutaba, "Virtual network embedding with coordinated node and link mapping", in: *INFOCOM 2009*, IEEE, Brazil, 2009.
- [8] J. Lischka, H. Karl, "A virtual network mapping algorithm based on subgraph isomorphism detection", in: *VISA '09: Proceedings of the 1st ACM workshop on Virtualized infrastructure systems and architectures*, Spain, 2009.
- [9] J. F. Botero, X. Hesselbach, A. Fischer, H. De Meer, "Optimal mapping of virtual networks with hidden hops", *Telecommunication Systems* 52 2013.
- [10] I. Houidi, W. Louati, W. B. Ameer, D. Zeghlache, "Virtual network provisioning across multiple substrate networks", *Computer Networks* 55 (2011) 1011 – 1023, special issue on Architectures and Protocols for the Future Internet, 2011.
- [11] M. Chowdhury, F. Samuel, R. Boutaba, "Polyvine: policy-based virtual network embedding across multiple domains" in: *Proceedings of the second ACM SIGCOMM workshop on Virtualized infrastructure systems and architectures*, VISA '10, India, 2010.
- [12] Y. Bejerano, Y. Breitbart, A. Orda, R. Rastogi, A. Sprintson, "Algorithms for computing QoS paths with restoration", *Networking*, IEEE/ACM Transactions on 13, 2005, pp.648–661.
- [13] N. Fujita, A. Iwata, "Adaptive and efficient multiple path pre-computation for QoS routing protocols", in: *Global Telecommunications Conference*, 2001. GLOBECOM '01. IEEE, USA, 2001.
- [14] F. Kuipers, P. Van Mieghem, T. Korkmaz, M. Krunz, "An overview of constraint-based path selection algorithms for QoS routing", *Communications Magazine*, IEEE 40, 2002, pp. 50–55.
- [15] L. Shen, M. Xu, K. Xu, Y. Cui, Y. Zhao, "Simple quality-of-service path first protocol and modeling analysis", in: *Communications*, IEEE International Conference on, China, 2004.
- [16] J. Rhoton "Cloud computing explained". Vol. 10. Breinigsville, PA USA: Recursive Press, 2010.
- [17] Walmara De Paula Herman and José Roberto de Almeida Amazonas. "Hop-by-hop routing convergence analysis based on paths algebra." *Electronics, Robotics and Automotive Mechanics Conference*, 2007. CERMA 2007. IEEE, 2007.
- [18] Azodolmolky, Siamak, Philipp Wieder, and Ramin Yahyapour. "Cloud computing networking: challenges and opportunities for innovations." *Communications Magazine*, IEEE 51.7, 2013.
- [19] A. Fischer, J. F. Botero, M. Duelli, D. Schlosser, X. Hesselbach, H. De Meer, "ALEVIN - a framework to develop, compare, and analyze virtual network embedding algorithms", *Electronic Communications of the EASST* 37, 2011.
- [20] J.F.Botero, X.Hesselbach "Greener Networking in a Network Virtualization Environment", *Computer Networks*, Volume 57, Issue 9, 19 June 2013, pp. 2021–2039.

Experimental evaluation of mobility management protocols for multimedia communications

J. Carmona-Murillo*, D. Cortés-Polo[†], J. Calle-Cancho[†], J. L. González-Sánchez[†], F. J. Rodríguez-Pérez*

* Department of Computing and Telematics System Engineering

University of Extremadura

Av. Universidad s/n, 1003, Cáceres, Spain.

jcarmur@unex.es; fjrodri@unex.es

[†] Research, Technological Innovation and Supercomputing Center of Extremadura (Cénits)

Carretera Nacional 521, km 41.8 10071, Cáceres, Spain

david.cortes@cenits.es; jesus.calle@cenits.es; joseluis.gonzalez@cenits.es

Resumen—The recent proliferation of multimedia portable devices, together with the increasing development of services and applications over mobile networks has made multimedia traffic one of the major challenges for the future mobile Internet. Current packet based mobile architectures such as 3GPP evolved packet system make use of all IP broadband services. Moreover, mobility management in 4G is made by means of host based and network based mobility management approaches. Mobile IPv6 and Proxy Mobile IPv6 are the most representative solutions standardized by the IETF. Recently, the IPv6 mobility support has been newly integrated into the kernel sources and Linux mobility ready kernels are available from versions 3.8.1. This support from IPv6 kernel space allows the development of open source mobility management protocols in the Linux user space. In this paper, we present an experimental study of Mobile IPv6 and Proxy MIPv6, based on their open source implementations. We analyze the handover latency characteristics in both protocols and provide both quantitative and qualitative performance measures of multimedia communications under different network conditions.

Palabras Clave—Mobile IPv6, Proxy Mobile IPv6, open source, experimental testbed, multimedia.

I. INTRODUCTION

Mobile data traffic in the Internet has experienced an exponential growth due to the widespread presence of multimedia capable mobile devices and the deployment of multiple wireless networks. The proliferation and variety of mobile applications and cloud services over the last few years is expected to intensify in the near future, specifically all those related to video over mobile networks [1].

In order to provide the continuous communication for mobile devices, the IETF (Internet Engineering Task Force) has standardized a fair number of mobility management protocols such as Mobile IPv6 (MIPv6) [2] or Proxy Mobile IPv6 (PMIPv6) [3]. These protocols are widely accepted as the most appropriate mechanisms for addressing seamless IP mobility in future wireless networks. In fact, the continuous development of these protocols has made the 3GPP (Third Generation Partnership Project) EPS (Evolved Packet System), commonly referred to as the 4G LTE (Long Term Evolution), to adopt PMIPv6 and Dual Stack MIPv6 for network based and host based mobility management respectively [4]. Mobile IPv6 is probably the most widely known IP mobility support protocol and it provides the mobility functionalities for a Mobile Node (MN) away from its home

network by updating its TCP/IP stack. This means that the MN is responsible for all mobility related signaling. Unlike the MIPv6 host based approach, in a network based mobility management approach such as PMIPv6, the serving network handles the mobility management on behalf of the MN; thus, the MN is not required to participate in any mobility related signaling.

The design of these mobility management protocols, similar to almost all protocols in use on the Internet, has followed an open, all inclusive process in which all documents are freely available over the Internet. Particularly, the IETF's open document process is a case study in the potential of the Open Source movement. The design process of an Internet protocol at IETF is an open call in which researchers, companies and individuals can participate [5]. Before its publication, the documents follow an extensive review process and only the standards that meet specific real world requirements and operate as expected, become true standards. In this process, open source implementation plays an important role due to its availability to the community. The experimentation in real environments allows the evaluation of the behavior and performance of the implementations under certain conditions. Thus, both the protocol design and implementation process share the same open philosophy and serve as a solid base for discussion and experimentation.

Moreover, the Linux based mobile platforms market is increasing rapidly and various open source solutions that offer seamless IP mobility in a device equipped with a mobility ready Linux kernel have been developed. In this paper, we present an experimental study of the handover latency characteristics of MIPv6 and PMIPv6, based on the open source implementations of these protocols (UMIP [6] and Open Air Interface PMIPv6 [7] respectively) in the Linux TCP/IPv6 stack.

This experimentation also includes a quantitative and qualitative study of multimedia communications during the movements of mobile users that indicate the behavior of the MIPv6 and PMIPv6 implementations and the user experience for real-time and streaming multimedia applications. The results have been obtained in a real scenario with real implementations of the protocols in Linux based devices and open source tools have also been used in order to generate, analyze and measure the mobile data traffic. Moreover, Netem network emulator

has been used in order to emulate the characteristics of the end to end path in the access network. This way, a wide variety of conditions such as delay, jitter and packet loss have been tested

The rest of the article is organized as follows. In Section II, we provide a background about mobility management and its most recent open source implementations. Then we describe the real infrastructure over which the experimental study has been done. Experimental results of handover latency and multimedia results are shown in Section III. Finally, Section IV concludes the paper.

II. RELATED WORK

In this section we describe briefly the most representative approaches such as MIPv6 and PMIPv6, depicted in Fig. 1 and Fig. 2 respectively, and give a review of their more recent open source implementations for the Linux IPv6 stack.

A. Mobile IPv6

Until now, Mobile IPv6 is the most representative mobile management scheme developed by the IETF on the way towards next generation mobile networks. It allows nodes to remain reachable while moving around in IPv6 networks. Without specific support for mobility, packets destined to a mobile node would not be able to reach it while the mobile node is away from its home link. In order to continue communication in spite of its movement, a mobile node could change its IP address each time it moves to a new link, but the mobile node would then not be able to maintain transport and higher-layer connections when it changes location.

The basic operation of mobile IP is as follows. The Mobile Node (MN) establishes a connection with the Correspondent Node (CN). A Home Agent (HA) serves as the anchor node in the Home Network that tracks the network connection point (location) of a user as the user moves around. Whenever a user changes their point of attachment to the network, the user registers with the HA, through Binding Update (BU) messages, informing of the user's current location and establishing a tunnel (IP in IP) between the HA and the MN located in a visited network. As soon as the MN receives a Routing Advertisement (RA) message from the new Access

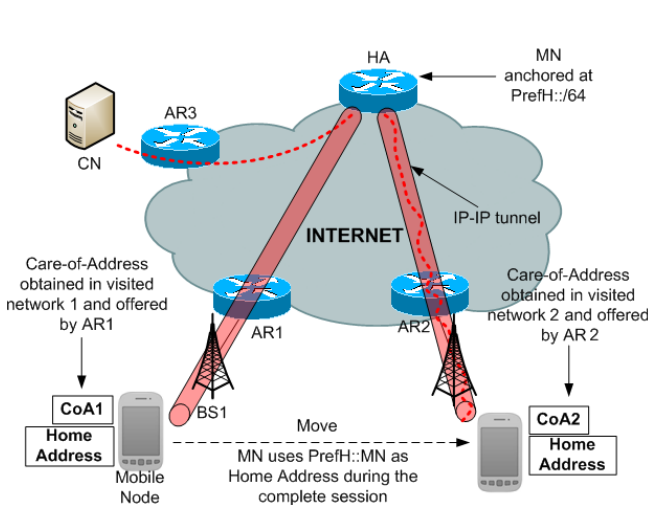


Fig. 1. Overview of MIPv6

Router (AR) in a foreign network, it obtains a new address called Care of Address (CoA) that is topologically correct for that network. The HA is the critical part of the system since it is on the path of both signaling and data for mobile users.

B. Proxy Mobile IPv6

The fundamental foundation of PMIPv6 is based on MIPv6 in the sense that it extends MIPv6 signaling and reuses many concepts such as HA functionality [8]. The new principal functional entities of PMIPv6 are the mobile access gateway (MAG) and local mobility anchor (LMA). The MAG typically runs on the AR that serves the MN. The main role of the MAG is to detect the MN's movements and initiate mobility related signaling with the LMA on behalf of the MN. In addition, the MAG establishes a tunnel with the LMA, enabling the MN to use an address from its home network prefix and emulate the MN's home network on the access network for each MN.

On the other hand, the LMA is similar to the HA in MIPv6. Its main role is to maintain reachability to the MN's address while it moves, and to store information necessary to associate an MN with its serving MAG. It also enables the relationship between the MAG and LMA to be maintained. As we can see in Fig. 1 and Fig. 2, both the MIPv6 and PMIPv6 have similar agents and messages. The main difference between them is that in MIPv6, the tunnel is established between the HA and the MN. So, the IP layer of the MN must be updated in order to operate with the MIPv6 protocol. However, in PMIPv6, the tunnel is created between the MAG and the LMA. This means that the mobile node is not aware of the mobility process and it is not necessary to update its TCP/IP stack.

C. Open source implementations

As well as the development of the RFC standards of IP mobility management protocols, where PMIPv6 was designed based on MIPv6, the implementations of both protocols has followed the same timeline and PMIPv6 implementation has been developed based on the MIPv6 code.

Linux implementation of mobility is divided into both kernel space and user space. This implementation strategy has been made in order to allow easily the extension of other

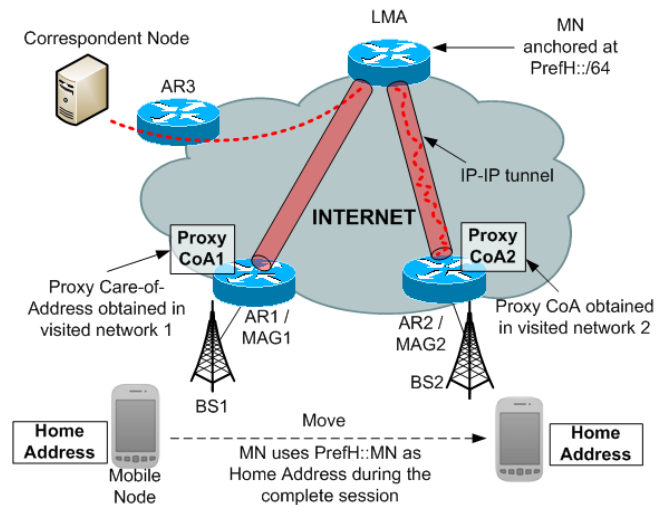


Fig. 2. Overview of PMIPv6

IP mobility protocols with a minimal kernel side and a user space where different protocols can be implemented. This way, the changes in the kernel can be kept to a minimum, which means more robust implementations. The kernel side support consists of a single module (`mip6.ko`) that has been newly integrated into the kernel sources since version 3.8.2 and most mobility protocol functionality such as MIPv6 or PMIPv6 are implemented in the user space.

In this work, MIPv6 support is provided by UMIP under a Linux mobility ready kernel. UMIP is an open source implementation of the MIPv6 protocol for the Linux operating system. To support PMIPv6, the Open Air Interface (OAI) PMIPv6 was developed extending UMIP functionality in order to support all necessary PMIPv6 messages and events. In both cases, the user space implements a daemon that takes care of the MIPv6 or PMIPv6 logic, such as tunnelling, binding signaling, security associations and IPv6 extensions, if needed. Some of these features are configured in each agent by means of configuration files that the daemon reads.

III. PERFORMANCE EVALUATION

A. Experimental setup and design

This section describes the setup and design of the MIPv6 and PMIPv6 scenarios used in the experimental study presented in this article. Fig. 3 illustrates the modules of the mobility agents involved in both approaches from an architectural point of view. The bottom layer represents the hardware and the wireless infrastructure; the Operative System (OS) layer includes the kernel space and the module to provide the mobility mentioned previously. In these experimental configurations, Ubuntu 12.04 OS is running in the mobility agents. The Linux 3.8.2 kernel with the mobility support has been compiled and installed in the OS to provide the basis to execute the mobility management daemons.

As can be observed in Fig. 3, the two open source approaches are implemented in the Linux user space. MIPv6 approach implements only UMIP as well as the MIPv6

daemon to provide the mobility management in the MN and the HA. The HA also requires the `radvd` daemon to send the Router Advertisement in the Home Network of the MN. The HA is integrated as a router in the access network, so it must exchange the necessary IPv6 routing information with other routers in the network. This function is implemented with Quagga routing software suite. This suite provides the implementation of Routing Information Protocol (RIP) for IPv6 using `zebra` and `ripngd` daemons.

For its part, PMIPv6 approach requires the OAI Proxy Mobile IPv6 implementation, in which the PMIPv6 functionality is included. Similarly to UMIP, OAI PMIPv6 provides a daemon that controls the protocol operation in the network entities (LMA and MAG). This daemon is implemented over UMIP and takes advantage of the similarities of both protocols. The LMA and MAG agents must also implement Free Radius server and client respectively to manage the IPv6 addresses and also implement the security mechanism to register the users in the access network.

Both LMA and MAG entities are located at the edges of the Proxy IPv6 domain and must implement the Quagga¹ routing software suite in order to exchange the routing information with other access network routers. Finally, The MAG and the access point must implement a `Syslog` server and client respectively to track the movement of the MN in the wireless network and to signal the movements to the LMA using the PMIPv6 Binding Update/ACK messages.

Fig. 4 depicts the Mobile IPv6 scenario used for the tests made in the experimental study. The MIPv6 scenario consists of several Cisco 1921/K9 routers which support IPv6 and RIPng that interconnect the HA, and the access points which provide the wireless access to the MN. These APs are Cisco Aironet 1130 AG series that support IEEE 802.11b/g specifications.

The PMIPv6 scenario (see Fig. 5) is similar to the previous MIPv6 one with the difference that in PMIPv6 two new agents

¹Quagga Routing Suite - <http://www.nongnu.org/quagga/>

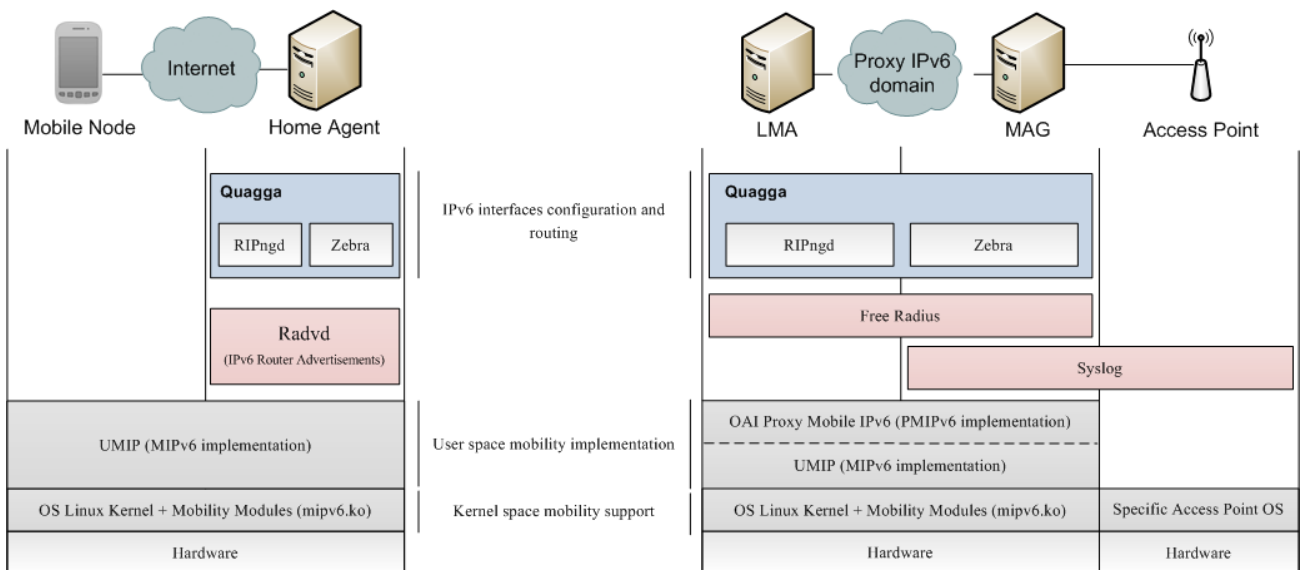


Fig. 3. Architecture of the mobility agents in MIPv6 and PMIPv6 approaches

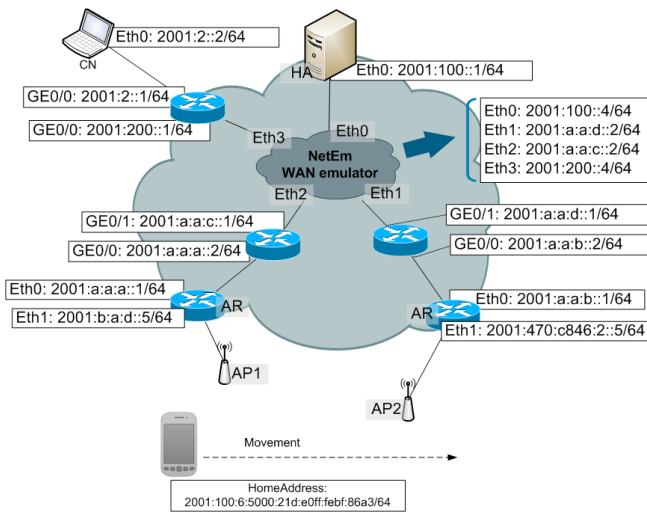


Fig. 4. Mobile IPv6 testbed topology

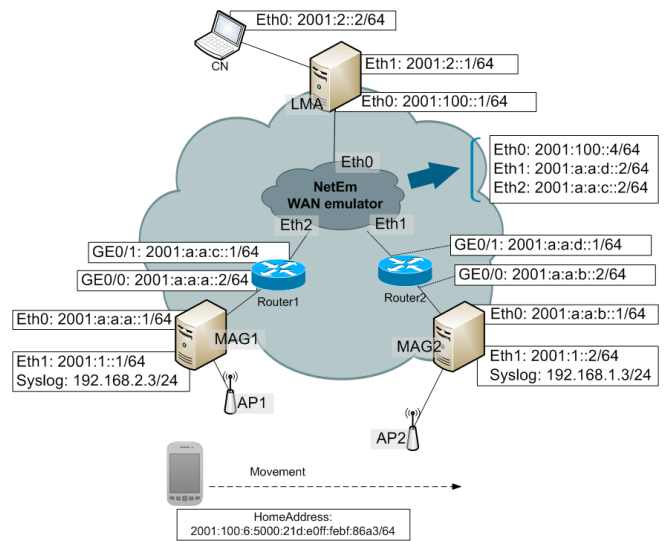


Fig. 5. Proxy Mobile IPv6 testbed topology

(LMA and MAGs) must be configured with the OAI PMIPv6 implementation. LMA functionality is similar to HA whereas MAGs are introduced instead of the access routers that serve the MN. In PMIPv6 the MN does not need any additional configuration to its default IPv6 stack. One important point is the requirement to emulate the conditions of the access network in order to evaluate how the access network characteristics affect mobile communication. For this purpose, the Netem network emulator is used. Netem allows a single Linux PC, set up as a router, to emulate a wide variety of network conditions (e.g. latency, jitter, packet loss,...). Both scenarios are similar so that the experiments obtain comparable results. The MN's handover is performed between two WLAN cells. These cells have enough overlapping surface so there is no possibility of the MN being unable to communicate with either of them. Notice that cell overlap is a requirement for seamless handover. Each WLAN cell belongs to a different IPv6 subnet.

B. Quantitative tests

This section provides a description of the test performed, as well as the open source tools used and the measurements done. The goal of these quantitative experimentations is to evaluate the performance of both MIPv6 and PMIPv6 Linux based open source implementations in terms of handover delay and the behavior of multimedia traffic under different circumstances such as packet delay or packet loss using the topologies observed in Fig. 4 and Fig. 5.

The handover latency (disruption time) is defined as the time elapsed between the last data frame being transmitted/received through the old interface and the first data frame being transmitted/received through the new interface. This time is a common parameter evaluated in a mobility scenario because the handover is one of the most critical processes. There are several comparative, analytical or simulation based studies of the behavior of handover latency in different mobility management protocols, however we have evaluated and compared the performance of both open source implementations using real infrastructure. The data given in this work allows to measure the gap between analytical or

simulated data and real implementations. We have evaluated this parameter by modifying the routing advertisement interval from 0.5 to 4 seconds.

The measurement of the handover latency consisted in sending packets at a high Constant Bit Rate (CBR), so that we could measure the time between the last packet received by the MN before the handover and the first packet received after the movement. The open source tool used to generate CBR traffic was *Ostinato*², whereas *Wireshark*³ was used to analyze the time stamped CBR traffic received by the MN during its movement. The handover delay for each of the RA intervals in MIPv6 is shown in Fig. 6, while Table I represents the numerical values of the handover delay parameter for the MIPv6 and PMIPv6 scenarios with a confidence interval of 95% for the different router advertisement values. In order to assure the confidence of the results, for each RA value 100 repetitions of the test have been made.

From the results shown in Fig. 6 we can quantify the dependency of the MIPv6 latency with the different Router Advertisement intervals, from 0.5 to 4 seconds. The minimum handover latency has been obtained with a RA interval of 0.5 and is 2.02 sec. The confidence interval for the mean at 95% limits is short as we can observe in Table I. As the RA interval increase, the confidence interval also increases due to the variability of the arrival to the MN of the unsolicited RA from the access router. As can be seen, PMIPv6 latency has not been included in Fig. 6. It is necessary to note that PMIPv6 is a network based approach and the home network is responsible for detecting that a new MN has been attached. In PMIPv6, the movement detection mechanism is not dependent on the RA messages. Table I demonstrates that the open source implementation of the protocol follows this behavior. The result obtained of the mean handover delay value for the 800 repetitions is, 2.788 seconds. The 95% confidence interval for this mean ranges from 2.788 to 2.790.

²Ostinato is an open-source, cross-platform network packet/traffic generator and analyser - <http://ostinato.org>

³Wireshark is a network protocol analyser for Unix and Windows - <https://www.wireshark.org>

Table I
VALUES OF THE HANDOVER LATENCY INTERVALS WITH A CONFIDENCE INTERVAL OF 95%

RA interval (s)	Hand. interval in MIPv6(s)		Hand. interval in PMIPv6(s)	
	Min.	Max.	Min.	Max.
RA interval = 0.5	2.020	2.149	2.788	2.788
RA interval = 1	2.287	2.477	2.789	2.790
RA interval = 1.5	2.449	2.697	2.779	2.780
RA interval = 2	2.810	3.055	2.783	2.784
RA interval = 2.5	2.944	3.218	2.781	2.784
RA interval = 3	3.252	3.624	2.794	2.797
RA interval = 3.5	3.485	3.864	2.807	2.811
RA interval = 4	3.540	4.078	2.790	2.790

As we previously mentioned, we have also evaluated the effect caused by IP mobility in multimedia communications. In this case, two types of multimedia transmissions have been considered and compared under different conditions. On the one hand, UDP real time communications based on RTP (Real Time Protocol) and, on the other hand, TCP multimedia streaming. Some network situations can affect the UDP multimedia stream such as network congestion, packet loss during handover and RTP packets arriving out of the playout time. In the TCP streaming case, the multimedia server sends the stream flow content to the receiver that is buffered in the client side. This mechanism avoids packet loss and minimizes the delay and jitter effects in the stream flow.

For each type of multimedia traffic, several parameters have been modified in the MIPv6 access network or in the PMIPv6 domain in order to evaluate its effect on the overall performance of the communication. These parameters are delay and packet loss, which were introduced by Netem emulator.

The experiments consist of the same movement of a terminal as in our testbed (causing a handover) during a real time or a streaming multimedia transmission under the different aforementioned conditions. The multimedia transmission is the Open Source Film Big Buck Bunny (duration, 120 s., bitrate approximately: 1626 kbps, resolution: 640x360). A

handover is performed in each test at second 53 approximately. In order to measure the transmission and to evaluate the received video quality by the MN, the Peak Signal to Noise Ratio (PSNR) indicator has been used. PSNR is generally considered to be a reference benchmark for developing objective perceptual video quality assessment models [9]. In this case, the reference model required by PSNR is the video transmitted in the testbed without the influence of the mobility management protocol and without any access network parameter modification. This reference is compared with the video received by the MN during the test using one of the evaluated mobility management protocols and changing the access network parameters. The MSU Video Quality Measurement Tool ⁴ was used for the PSNR evaluation.

Fig. 7, Fig. 8 and Fig. 9 shows a summary of the multimedia experimentation results. Fig. 7, and Fig. 8 show the behavior of both real time and streaming multimedia communications respectively in MIPv6. In this case, just three representative tests are present in the graph. The first one is a communication without any access network parameter modification, the access network in the second one introduces a delay of 100 ms, and finally, a test with a packet loss of 10 %. The result of the PSNR evaluation of all the experiments is presented in Fig. 9.

⁴MSU Video Quality Measurement Tool is a program for objective video quality assessment. It provides functionality for both full-reference (two videos are examined) and single-reference (one video is analyzed) comparisons- http://compression.ru/video/quality_measure/video_measurement_tool_en.html

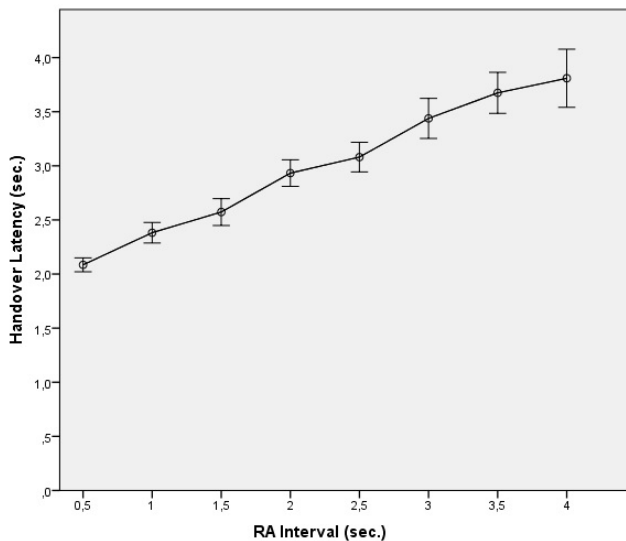


Fig. 6. Handover latency in MIPv6 experiments

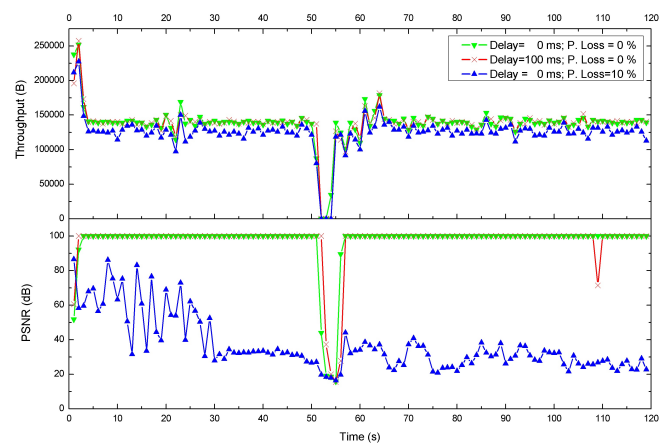


Fig. 7. Throughput and PSNR in real time experiments

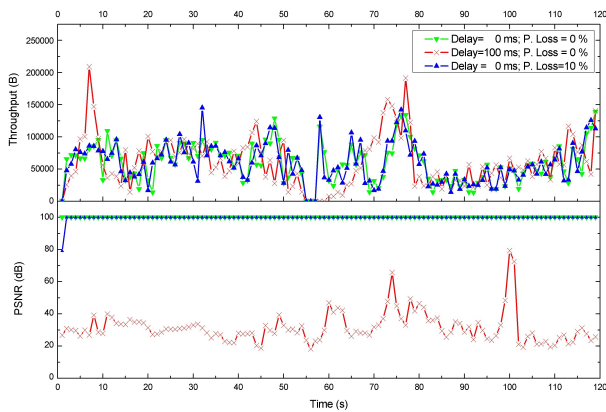


Fig. 8. Throughput and PSNR in TCP streaming experiments

Fig. 7 presents packet loss and delay influence in a real time communication transmitted in the MIPv6 testbed. The results show that the PSNR is extremely sensitive to packet loss in RTP. When this parameter is increased, many RTP packets are lost and the video frames must be predicted using previous frames. This prediction reduces the PSNR value of the experiment. In this figure, a high packet loss rate can also be observed and the throughput of the communication decreases compared with the delay test.

In Fig. 8 we can observe a different behavior by the PSNR. In this case, three TCP streaming communications are compared in the same conditions as stated before. The streaming application provides a buffer that avoids packet loss and reduces the delay and the jitter of the packets. As can be observed in Fig. 8, when the delay is high the application must wait to recompose the video transmission and the PSNR is penalized due to long wait times. In contrast, when the packet loss is high, the video frames are retransmitted by the transmitter and the PSNR value is higher compared with the real time transmission. In this case there are no wait times and the PSNR is not penalized. This implies that PSNR parameter in streaming transmission decrease when the delay of the access network increases. As could be observed, although the handover is produced, the PSNR is not affected because of the buffer. Only when the delay is increased, the PSNR decreases.

Fig. 9 shows the different experiments made in both MIPv6 and PMIPv6 testbed using UDP real time and TCP streaming communications. In each of these four categories, eight tests have been made and each one represents a PSNR value in the plot (5 repetitions have been made for each PSNR value). Regular case refers to tests made with the configuration of Netem at Delay = 0 and without packet loss. The behavior of MIPv6 corresponds to the one explained previously in Fig. 9 and Fig. 8 whereas in PMIPv6 we can observe some differences. Real time transmission in the PMIPv6 testbed demonstrates the same behavior as in MIPv6 when the delay of the access network increases. When delay exceeds 75 ms, the PSNR decreases because the packets arrive too late to the MN and the video frames are discarded by the player as the time limit to be reproduced has been exceeded. In our experiments, when the handover occurs, the serving MAG of the visited network introduces long delays to forward packets to the MN. These latencies penalize the communication and

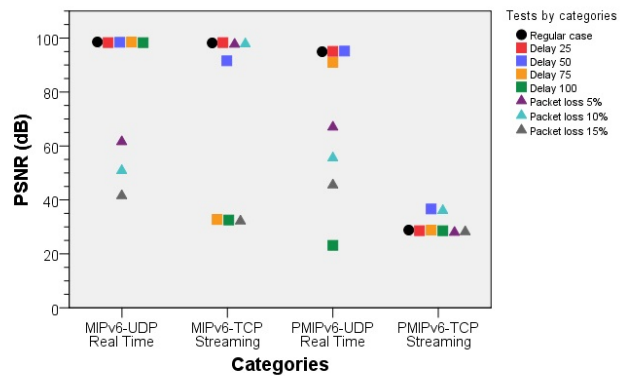


Fig. 9. Summarize of the obtained PSNR results in the experiments conducted

the PSNR values. The same effect is produced in PMIPv6 streaming, where the PSNR achieves low values because the reproduction pauses until it has enough information to continue the reproduction due to buffer starvation. Despite the PSNR values, that reflect the comparison between the original video and the received one, the videos seems to be reproduced at good quality. In the next section, a user experience valuation is presented in order to compare qualitative versus quantitative results.

C. Qualitative tests

In the previous section, some quantitative tests and experimental results have been presented. The quantitative tests offer a measurement of various well known parameters which affect communication. In contrast, the qualitative tests deal with user expectation, satisfaction and overall experience. A typical user related measure is the Mean Opinion Score (MOS) [10], which can be determined from subjective ratings by real users. With the purpose of reflecting how users perceive multimedia content in mobility scenarios managed by open source implementations of MIPv6 and PMIPv6, in this section we show the results obtained from user valuations. In total, 30 real users were asked to evaluate the perceived quality according to the five point MOS scale (MOS scores goes from 1 to 5, 5 being the best). Each test was conducted in a similar way as the one presented in the previous section. In this case, Netem is configured in the access network in order to introduce different delays (0, 25, 50 and 100 ms). Fig. 10 shows the user valuation of both real time and streaming transmissions under different delay conditions in the MIPv6 and PMIPv6 experimental testbed.

Fig. 10 (a) shows the results obtained to transmit UDP real time content. As can be observed, the MOS rate decreases when the delay of the access network increases. With the introduction of delay, some packets arrive to the MN out of the reproduction timestamp and they are discarded producing image pixelation and a gap between frames. As can be observed in the figure, the users penalize these effects.

The TCP streaming transmission increases the overall MOS score in both management protocols because all packets are stored in a buffer avoiding image pixelation, and a gap in the video reproduction produced by packet loss or packet discard.

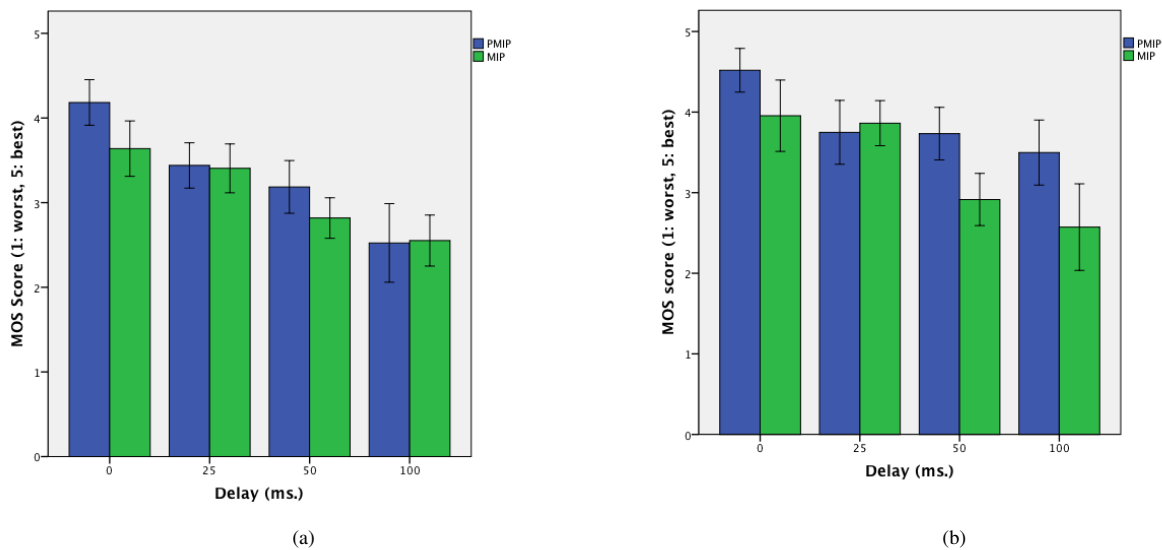


Fig. 10. Mean Opinion Score of (a) real time and (b) TCP streaming transmissions with a confidence of 95%

The inconvenience introduced by TCP streaming is the long wait time produced by the buffer starvation. This effect can be observed when the delay of the access network increases. The handover latency is increased and this means that the wait time until the buffer is filled with enough packets to continue the reproduction is also increased. This effect is shown in Fig. 10 (b).

IV. CONCLUSIONS

Testbed are, by far, the most realistic method for evaluating performance, as they practically use the protocol implementations and the hardware that is the same, or very similar, to the one used for the production networks. However, large testbed are expensive to build and manage and for very large (or highly mobile) networks, practically impossible to implement. Furthermore, the degrees of freedom in a testbed is significantly reduced in comparison with a mathematical analysis or network simulation. Assuming these limitations, in this article we have conducted a study related to the performance of real IPv6 mobility management implementations in the IPv6 Linux stack, evaluating the current state of these protocols by means of an experimental testbed.

We have focused on the most representative IETF solutions: MIPv6 and PMIPv6. Qualitative and quantitative results have been obtained from both real time and streaming multimedia applications under different network conditions. Quantitative results have shown the handover latency produced by both open source implementations. We quantify the dependency of the MIPv6 handover delay in the Routing Advertisement interval, whereas the PMIPv6 implementation is not dependent of RA messages, as expected. Moreover, the PSNR video quality indicator has been used to evaluate the multimedia transmission. It should be noted that streaming traffic in the PMIPv6 testbed give low PSNR values due to the short additional delays produced by the MAG which penalizes the PSNR value because some wait times are introduced due to buffer starvation.

Finally, in order to compare these quantitative results of the multimedia transmissions with a subjective valuation, user

perceptions are evaluated. In this case, streaming multimedia provides a better valuation than real time due to penalization of the handover disruption time. A better user experience is achieved when handover latency is reduced.

REFERENCIAS

- [1] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019, White Paper, Cisco, February 2015.
- [2] C. Perkins, D. Johnson, and J. Arkko, "Mobility Support in IPv6," RFC 6275, IETF, July 2011. Available: www.ietf.org/rfc/rfc6275.txt.
- [3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6," RFC 5213 (Proposed Standard), IETF, August 2008, updated by RFC 6543. Available: www.ietf.org/rfc/rfc5213.txt.
- [4] J. C. Zuniga, C. J. Bernardos, A. De La Oliva, T. Melia, R. Costa, and A. Reznik, "Distributed mobility management: a standards landscape," *IEEE Communications Magazine*, 51(3):80-87, March 2013.
- [5] C. DiBona and S. Ockman. "Open sources: Voices from the open source revolution," O'Reilly, 1999.
- [6] UMIP. "Mobile IPv6 and NEMO for Linux," <http://umip.org>, May 2015.
- [7] EURECOM, "Open Air Interface Proxy Mobile IPv6 (OAI PMIPv6)," <http://www.openairinterface.org/openairinterface-proxy-mobile-ipv6-oai-pmipv6>, Aug. 2013.
- [8] Ki-Sik Kong, W. Lee, Y-H. Han, M-K Shin and H. You, "Mobility management for all-IP mobile networks: Mobile IPv6 vs. Proxy Mobile IPv6," *IEEE Wireless Communications*, vol. 15, no. 2, pp. 36-45, April 2008.
- [9] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication systems*, vol. 49, no. 1, pp 35-48, January 2012.
- [10] M. Fiedler, T. Hossfeld and P. Tran-Gia, "A Generic Quantitative Relationship between Quality of Experience and Quality of Service," *IEEE Network*, vol. 24, no. 2, pp. 36-41, March-April 2010.

QoS y QoE en Comunicación Multimedia (QQCM)

Optimización y securización de servicios multimedia interactivos: impacto en las relaciones comerciales online

Guillermo Azuara Guillén, Julián Fernández Navajas, José Ruiz Mas, José Luis Salazar Riaño, José María Saldaña Medina.

Grupo CeNITEQ– Instituto de Investigación en Ingeniería de Aragón

Dpt. IEC. EINA, Universidad de Zaragoza

Edif. Ada Byron, 50018, Zaragoza

{gazuara, navajas, jruiz, jsalazar, jsaldana }@unizar.es

Blanca Hernández Ortega, Julio Jiménez Martínez, María José Martín de Hoyos

Grupo de Investigación Generés

Dpt. Dirección de Marketing e Investigación de Mercados, Universidad de Zaragoza

c/ Gran Via, 2, 50005, Zaragoza

{bhernand, jjimenez, mjhoyos }@unizar.es

Resumen- A pesar del interés e importancia de los servicios multimedia interactivos, éstos apenas han sido incluidos en plataforma de servicios de las empresas, debido principalmente a que su desarrollo todavía adolece de ciertas limitaciones que les impide ofrecer la calidad deseada. Desde una perspectiva de red, estos servicios generan flujos extremo a extremo ineficientes en su transporte, más aun si éstos han de ser securizados. Su optimización requiere lograr un equilibrio entre eficiencia y el mantenimiento de la calidad y de la seguridad dentro de unos parámetros aceptables para el usuario. En el ámbito económico-empresarial, el usuario suele percibir un elevado riesgo y una falta de seguridad relacionada con el empleo de estos servicios, lo cual tiene un impacto negativo en su confianza, en su relación con la empresa y en su toma de decisiones. Para cubrir estas limitaciones, surge la necesidad de aplicar un enfoque interdisciplinar que complementa ambas visiones de trabajo.

Palabras Clave- servicios multimedia interactivos, relaciones comerciales online, QoS, valor para el usuario

I. INTRODUCCIÓN

Los servicios multimedia interactivos están adquiriendo un interés e importancia crecientes. Sin embargo, apenas han sido incluidos en plataforma de servicios de las empresas, debido principalmente a que su desarrollo todavía adolece de ciertas limitaciones que les impide ofrecer la calidad deseada. Desde la perspectiva de la red, estos servicios suelen presentar una eficiencia muy baja, siendo denominados en ocasiones “servicios de paquetes pequeños” (*small-packet services*) [1]. Sus estrictos requisitos temporales conllevan que la información deba ser fragmentada en paquetes de pocos bytes, enviados a altas tasas. Esto trae consigo un *overhead* elevado, debido a la relación existente entre el tamaño de las cabeceras y el de la información útil (*payload*).

Para paliar esta ineficiencia, en los últimos años nuestro grupo de investigación ha propuesto un mecanismo de optimización de tráfico denominado TCM (*Tunneling Compressing and Multiplexing*), para su estandarización a nivel global en el IETF (*Internet Engineering Task Force*). Este mecanismo de optimización combina la compresión de

cabeceras, la multiplexión y el tunelado, cuando un número de flujos de tiempo real comparten un mismo camino en la red. De esta manera, los paquetes pueden viajar comprimidos extremo a extremo, y el *overhead* causado por las cabeceras se reparte, logrando así ahorros efectivos de ancho de banda.

Sin embargo, el mecanismo propuesto requiere, como contrapartida al ahorro obtenido, un pequeño retardo necesario para juntar un número suficiente de paquetes y, por otro lado, otro retardo de procesamiento para comprimir las cabeceras. Para que la optimización no reduzca la calidad percibida por el usuario, se deben definir unos límites para estos retardos adicionales.

Por otra parte, si se desea añadir seguridad a los servicios interactivos (por ejemplo, cifrando o firmando la información), su eficiencia se reduce todavía más, debido a la necesidad de incluir cabeceras de seguridad adicionales. El método de optimización propuesto no contempla de momento la securización de los flujos a optimizar, pero diferentes mecanismos basados en la delegación de confianza podrían permitir la integración de la seguridad y abrir simultáneamente el camino hacia la mejora de la eficiencia: podemos pensar, por ejemplo, en la posibilidad de securizar conjuntamente un número de flujos que comparten un camino en la red. De esta manera, el *overhead* correspondiente a las cabeceras de seguridad se repartiría entre todos los flujos.

Por lo tanto, podemos concluir que la optimización del tráfico de servicios interactivos requiere lograr un equilibrio entre el incremento de la eficiencia de la red y el mantenimiento de la calidad y de la seguridad dentro de unos parámetros aceptables para el usuario. En la actualidad se está investigando en la definición del mecanismo de optimización de tráfico y en el establecimiento de los límites del retardo, para mantener la calidad subjetiva. Posteriormente se analizarán las sinergias entre la optimización de tráfico y la incorporación de seguridad que permitan proporcionar al usuario servicios de tiempo real con calidad y seguridad en el ámbito comercial.

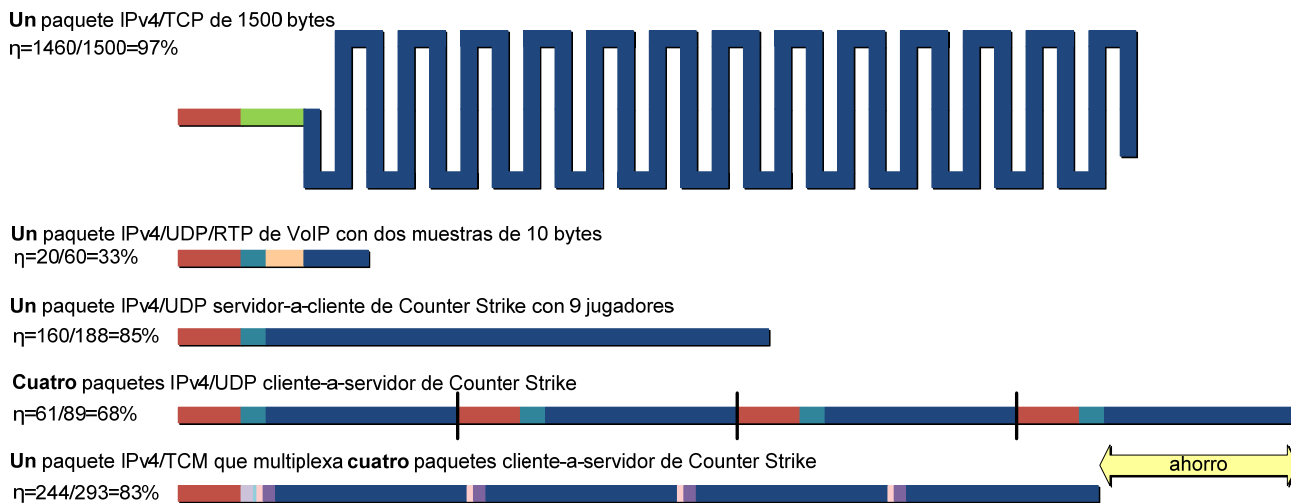


Fig. 1. Ejemplos de eficiencia de paquetes de Internet para diferentes servicios

II. ANTECEDENTES

A. TCM (Tunneling, Compressing and Multiplexing) para optimización de tráfico Internet.

En los últimos años el desarrollo de Internet está favoreciendo el despliegue de diferentes servicios interactivos en tiempo real, así como su inclusión en plataforma de servicios de las empresas, lo cual permite establecer formas más sofisticadas de comunicación entre los interlocutores. Todo esto conlleva que los usuarios, y el público en general, adquieran un mayor protagonismo, ya que dichos servicios no sólo posibilitan la interacción entre personas, sino que también potencian su participación activa (por ejemplo, mediante una conversación o una videoconferencia con el personal de la empresa o con otros usuarios en la propia plataforma). Este mayor protagonismo implica un mayor poder del cliente en las relaciones comerciales, siendo la confianza y seguridad elementos clave a la hora de interactuar con otros actores en el mundo virtual.

Los primeros servicios interactivos ampliamente desplegados en Internet fueron la Voz sobre IP (VoIP) y la videoconferencia, que usaban la red para establecer comunicaciones similares a las proporcionadas por las redes tradicionales de conmutación de circuitos. Al mismo tiempo, el éxito de otros servicios más recientes, como los juegos *online* o la televisión IP, ha contribuido al aumento del tráfico global con requerimientos estrictos de tiempo real. En este sentido, estamos asistiendo también a una proliferación de contenidos generados por el usuario con nuevos servicios emergentes para su distribución, que están empezando a incluirse en plataforma de servicios de las empresas.

El diseño inicial de Internet no contemplaba la posibilidad de entregar la información cumpliendo requisitos temporales. Así, la red se diseñó con el objetivo de enviar paquetes grandes, pues cada unidad de información debe disponer de una serie de cabeceras de distintos niveles para poder llegar a su destino. Cada tecnología tiene un tamaño máximo de paquete, y si se envían paquetes de ese tamaño o algo inferiores, el coste de las cabeceras (*overhead*) es

asumible, al repartirse entre un gran número de bytes de información. Sin embargo, si se desea utilizar Internet para soportar servicios en tiempo real, es necesario enviar actualizaciones frecuentes de información y esto obliga a que los paquetes sean pequeños, pues los requerimientos temporales impiden esperar con el fin de acumular más bytes de información. Esto hace que en el caso de servicios en tiempo real, la eficiencia puede llegar a ser realmente baja. La Fig. 1 muestra el contraste entre la alta eficiencia de un paquete de 1500 bytes, y la baja eficiencia de uno de VoIP o de un juego online. Sin embargo, la optimización puede paliar esta ineficiencia.

En esta situación, los operadores de red están observando un cambio en la distribución agregada del tráfico que deben transportar sus redes: la proporción de flujos de paquetes pequeños aumenta [1]. Por ello, se hace necesario definir nuevos mecanismos de red que permitan adaptarse dinámicamente a las variaciones de tráfico, con el fin de mejorar los parámetros de calidad. Ello conlleva que los diferentes actores presentes en el escenario cooperen entre sí para adaptar las infraestructuras tradicionales de red a estos nuevos patrones de tráfico, mientras se garantizan los requisitos de calidad -ancho de banda, retardo, seguridad, etc.- que se traduzcan en una calidad de la experiencia (*Quality of Experience, QoE*) apropiada a los usuarios finales. Además se presenta un problema adicional, ya que estos flujos tienen en ocasiones un comportamiento imprevisible; por ejemplo, la mensajería instantánea presenta picos de tráfico difíciles de prever en determinados momentos y lugares donde coinciden a la vez muchas personas (eventos deportivos, exposiciones), o en el lanzamiento de un nuevo producto para su venta en una plataforma empresarial.

Vemos, por tanto, que estos servicios generan flujos extremo a extremo de larga duración, con baja eficiencia en el transporte y que comparten un mismo camino de red con otros flujos. Así pues, la optimización de estos flujos mejoraría la eficiencia de la red, consiguiendo ahorros efectivos de ancho de banda y la disminución del número de paquetes por segundo (pps) que la red debe gestionar y,

además, minimizaría el impacto del procesado y del gasto energético en los nodos de la red.

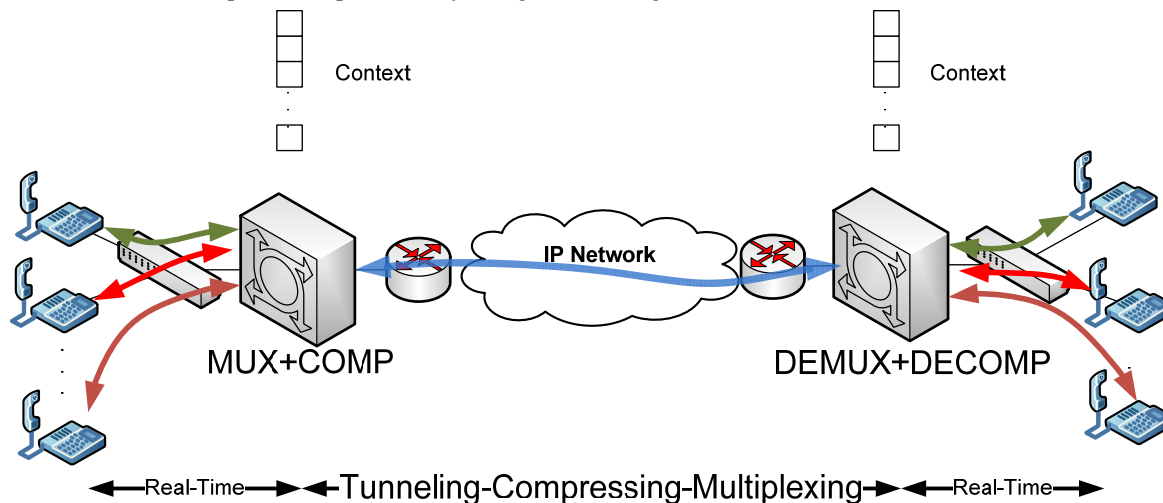


Fig. 2. Esquema de Tunneling, Compressing and Multiplexing Traffic Flows

Para reducir el *overhead* se han desarrollado algoritmos de compresión de cabeceras (RFC 3545, RFC 2507, RFC 5225), que aprovechan la redundancia de muchos de los campos, que son iguales en todos los paquetes de un flujo. El problema de estos algoritmos es que sólo pueden funcionar punto a punto, puesto que un paquete con una cabecera incompleta no puede viajar por Internet. Para enviarlos extremo a extremo debe usarse un protocolo de tunelado (p. ej. RFC 2661), que vuelve a añadir gran parte del *overhead* que queríamos evitar. Pero si la compresión de cabeceras se combina con la multiplexión (RFC 3153) de varios paquetes dentro de uno más grande, entonces el *overhead* causado por el túnel se reparte, y pueden conseguirse ahorros de ancho de banda mucho mayores.

El esquema propuesto puede observarse en la Fig.2, donde la optimización del tráfico se consigue mediante la combinación de compresión de cabeceras, multiplexión y tunelado (TCM).

En esta misma línea, en 2005 el IETF estandarizó “Tunneling Multiplexed Compressed RTP” (TCRTP, RFC 4170), que combina el método de compresión “Enhanced Compressed RTP” (ECRTP, RFC 3545) con el protocolo de multiplexión “Point to Point Protocol Multiplexing” (PPPMux, RFC 3153) y que finalmente envía los paquetes con un túnel “Layer 2 Tunneling Protocol” (L2TP, RFC 2661), permitiendo así ahorros significativos (hasta un 50% del ancho de banda) para flujos de VoIP [2]. Sin embargo, esta solución está circunscrita a flujos que usen RTP/UDP/IP, y no existe un estándar que permita esta misma optimización de tráfico para otros flujos de paquetes pequeños. Si acaso podemos encontrar en la literatura soluciones ad hoc a esta problemática y para casos y tecnologías muy concretas, como [3], que también utiliza PPPMux para multiplexar tráfico.

Como contrapartida a la optimización, el método de multiplexión-compresión añade unos pequeños retardos causados por el procesado y por el tiempo requerido para reunir el número de paquetes necesarios. Diferentes trabajos [2], [4], [5] han mostrado que estos retardos adicionales se pueden mantener por debajo de un umbral, de forma que la calidad subjetiva del usuario final no se vea afectada.

B. Seguridad para los flujos de comunicaciones optimizados

La inclusión de mecanismos de seguridad en las comunicaciones digitales, y más concretamente en Internet, se torna esencial debido tanto a la necesidad de salvaguardar la información intercambiada (cifrando las comunicaciones) como la de verificar la identidad de los emisores de las mismas (mediante mecanismos de firma digital). Para añadir esta seguridad en las comunicaciones se han desarrollado diversos protocolos que actúan a distintos niveles de la pila de protocolos: *Encryption Control Protocol* (ECP, RFC 1968), *IPSec* (RFC 4301), *Transport Layer Security* (TLS, RFC 5246), etc., cuya seguridad reside en una adecuada gestión de la confianza de las claves de los usuarios. Debemos tener en cuenta, por tanto, un planteamiento estratégico de esta confianza y una implementación táctica de los diversos protocolos criptográficos.

Las infraestructuras de clave pública (*Public Key Infrastructure*, PKI) permiten a cada usuario gestionar autónomamente sus claves certificadas por autoridades confiables. Una alternativa que no tuvo excesiva repercusión fue la *web of trust* [6], en la que cada usuario debía ganar la confianza de los miembros de la red. Otra posibilidad es el cifrado basado en identidad [7], donde la gestión de las claves se cede parcialmente a las autoridades confiables. En cuanto a la implementación táctica, el uso de seguridad en distintas capas se ajusta a la amplia diversidad de escenarios posibles, asegurando la(s) capa(s) con los operadores criptográficos y las claves más adecuadas para cada circunstancia.

Si centramos los objetivos en las capas en las que se implementa TCM, debemos ver cómo se gestionan los requisitos de seguridad a nivel de red y superiores, en beneficio de la eficiencia del servicio. De hecho, la seguridad a nivel de red se basa en el empleo de *IPsec*, que proporciona un conjunto de servicios básicos de seguridad y sustenta un marco de trabajo para futuros protocolos seguros. Los dos protocolos que ofrece *IPsec*, *IP Authentication Header* (AH) e *IP Encapsulating Security Payload* (ESP), permiten adaptar el nivel de seguridad a los requisitos de la comunicación cifrando y/o firmando los paquetes de datos, pudiendo ser usados tanto en modo transporte como en modo túnel.

De la misma manera que a nivel de red se encuentra IPsec, en la capa de transporte encontramos TLS, que proporciona también las opciones de autenticación y cifrado de la información en un servicio tan extendido como puede ser el "servicio web". De manera más específica, encontramos el protocolo *Secure RTP* (SRTP, RFC 3711), que actúa a nivel de sesión y permite securizar las comunicaciones en tiempo real (VoIP, videoconferencias, etc.) que son susceptibles de ser optimizadas con TCM.

C. Servicios interactivos en el ámbito económico-empresarial

En los últimos años, el desarrollo de las nuevas Tecnologías de la Información y Comunicación (TICs) y el incremento de la popularidad de medios sociales han favorecido la proliferación y la difusión de estos nuevos servicios interactivos, que fomentan la participación de diferentes agentes y generan nuevas formas de recoger, procesar y producir información [8]. El uso de estos servicios constituye una fuente de rápido crecimiento para las empresas, ya que generan nuevos modelos de negocio, permiten alcanzar soluciones óptimas en distintos ámbitos, como la ingeniería, la sanidad, la vivienda o el sector del automóvil, y pueden ser aplicados tanto en contextos industriales como en mercados de gran consumo [9], [10].

En el ámbito comercial, la inclusión y el empleo de estos servicios modifica las relaciones existentes entre clientes y entre el cliente y la empresa, facilitando la participación activa de todos los agentes involucrados en el proceso y modificando las reglas de juego vinculadas a este tipo de relaciones [11], [12]. De este modo, la posibilidad de acceder a un mayor volumen de información, de comunicarse en tiempo real, así como de conocer a la(s) persona(s) que se encuentra(n) al otro lado de la pantalla, aporta un valor a las relaciones comerciales, que no había sido alcanzado con anterioridad.

Las transacciones a través de Internet han dejado de ser intercambios diádicos empresa-cliente, en los que la primera controlaba la información y tomaba sus propias decisiones, pasando a adoptar una estructura dinámica de red donde diferentes agentes forman parte del proceso, integran recursos y participan en la relación [13], [14]. Estudios fidedignos estiman que los beneficios derivados del uso de este tipo de servicios alcanzarán los 500 millones de dólares para 2018 [15].

Si bien la implementación de servicios interactivos desde una perspectiva técnica puede suponer incrementos en la eficiencia y la productividad (p.ej. ahorro en costes y tiempo), todavía existen importantes retos sin resolver relacionados con la aceptación y uso por parte del usuario final [16], [17]. En primer lugar, debemos señalar que el usuario sigue percibiendo un elevado riesgo y una falta de seguridad relacionados con el empleo de estos servicios [18], [19], lo que tiene un impacto negativo en su confianza, en su relación con la empresa y en su toma de decisiones. En segundo lugar, hemos de destacar que la mayor parte de la investigación económico-empresarial parte de servicios ya desarrollados, por lo que los resultados alcanzados profundizan en los errores, limitaciones y peculiaridades de su empleo, proponiendo soluciones a posteriori, dirigidas a aquellas empresas que han decidido utilizar este tipo de servicios. Además, el objetivo de estas investigaciones no es modificar los servicios existentes, por lo que no pueden alcanzar en ningún caso conclusiones

relacionadas con el proceso de desarrollo y optimización del servicio, que debe llevarse a cabo antes de implementarse en una plataforma empresarial. En último lugar, debemos subrayar que la mayoría de los estudios realizados en el área económico-empresarial se han centrado en la interacción del usuario con la tecnología [8].

III. IDENTIFICACIÓN DE PROBLEMAS Y OBJETIVOS

A. Identificación de problemas

Como ya ha sido comentado y a modo de resumen se identifican dos problemas claramente diferenciados. Por un lado, a pesar del interés e importancia de los servicios multimedia interactivos, estos apenas se han incluido en plataforma de servicios de las empresas, debido principalmente a que su desarrollo todavía adolece de ciertas limitaciones que les impide ofrecer la calidad deseada.

Por otro lado, en cuanto a la aceptación y uso por parte del usuario final en un ámbito económico-empresarial, a día de hoy, no se ha profundizado en el impacto ejercido por el empleo de estos servicios en las interacciones personales que tienen lugar en entornos comerciales entre proveedor/empleado y cliente, entre diferentes proveedores de servicios y entre los propios clientes de los servicios [17].

Para cubrir estas limitaciones, surge la necesidad de aplicar una perspectiva interdisciplinar que, desde el punto de vista del usuario final, aborde cuestiones interpersonales vinculadas a su empleo en las relaciones comerciales. En definitiva buscamos la explotación de las sinergias entre la optimización de tráfico y la incorporación de seguridad que permitan proporcionar al usuario servicios de tiempo real con calidad y seguridad en el ámbito comercial.

Para ello se proponen los objetivos detallados en los dos siguientes apartados.

B. Caracterización, optimización y securización del tráfico de servicios interactivos que generen flujos de baja eficiencia y que puedan ser utilizados en plataformas comerciales online.

Este objetivo pretende profundizar y seguir avanzando en la definición y documentación de la propuesta de un estándar global de optimización de tráfico, centrándose en los servicios interactivos. Hasta la fecha se han dado pasos en esta línea, pero la inclusión de seguridad requiere un estudio pormenorizado de los escenarios, de las relaciones entre los diferentes actores y de los requerimientos temporales y de procesado. De esta forma, el estándar permitiría la optimización del tráfico y su securización, manteniendo los requerimientos de calidad subjetiva y proporcionando el nivel de seguridad adecuado para el cliente.

La inclusión de mecanismos de optimización en las redes de nueva generación se ve como necesaria para garantizar su flexibilidad, aunque apenas haya sido desarrollada hasta la actualidad. La alta movilidad de los usuarios hace más complicada la planificación de las redes. En efecto, puede ocurrir que un gran número de usuarios acudan a un evento público y generen un pico de tráfico para el que la red no está preparada. El método de optimización permite dotar a las redes de un nivel de flexibilidad, intercambiando procesado por ancho de banda: si la demanda de tráfico excede la capacidad de la red, la optimización puede

activarse dinámicamente permitiendo su operación incluso por encima de su capacidad planificada inicialmente.

No debemos perder de vista que este objetivo se centra en la sinergia entre optimización y seguridad. Efectivamente, la seguridad requiere cabeceras adicionales y esto acentúa la baja eficiencia de los servicios interactivos. Por eso, su optimización puede resultar especialmente beneficiosa. Pero deben tenerse en cuenta las relaciones de confianza entre los actores del escenario, por ejemplo, a la hora de cifrar conjuntamente un número de flujos de diferentes usuarios. Una confianza basada en los proveedores de servicios de comunicaciones podría mejorar el rendimiento de la red; sin embargo, este requisito no está implícito en la definición de proveedor de servicio. En cambio la confiabilidad es algo que sí define a la Autoridad de Certificación en un escenario clásico de Infraestructura de Clave Pública (PKI) y deberá tenerse en cuenta para dar solución al objetivo planteado.

C. Estudio desde un enfoque económico-empresarial de las consecuencias derivadas del empleo de servicios interactivos seguros en las relaciones comerciales.

Con este objetivo pretendemos evaluar la aplicabilidad de la optimización del tráfico y sus beneficios, midiendo el valor generado no sólo para las empresas de telecomunicaciones, que podrán optimizar su tráfico de red, sino también para el usuario final y para cualquier empresa que comercialice sus productos a través de la red. De esta forma añadimos a la evaluación clásica de calidad de experiencia un enfoque económico-empresarial que supone una ayuda en la toma de decisiones para mejorar los servicios escogidos. El punto de partida de este objetivo lo constituyen los clientes, considerados el principal activo de toda organización empresarial. La posibilidad de establecer nuevos mecanismos de interrelación, a través de los cuales los clientes tengan la posibilidad de adquirir mayor información y de relacionarse en tiempo real supone una fuente de valor fundamental para diferenciarse de la competencia y para generar rentas presentes y futuras.

Los servicios interactivos se analizarán en un contexto comercial caracterizado por la posibilidad de establecer diferentes interacciones sociales online entre todos los agentes involucrados. Así, estas interacciones fomentan el establecimiento de relaciones a largo plazo, y el desarrollo de comportamientos no transaccionales por parte del cliente que pueden constituir una fuente generadora de valor en el largo plazo. Esto supone un reto en la medida en que, por un lado, requiere un horizonte de trabajo ampliado y, por otro, conlleva conocer las interrelaciones sociales llevadas a cabo entre los clientes, puesto que los beneficios financieros provendrán de cómo los comportamientos de unos clientes afectan a las decisiones de otros.

Tal y como ha sido explicado anteriormente, las interacciones en tiempo real pueden generar relaciones a largo plazo entre los agentes involucrados en un contexto comercial. No obstante, estas interacciones tendrán efectos positivos siempre y cuando cumplan una serie de requisitos de seguridad que deben ser correctamente percibidos por los usuarios. Así, aquellos entornos y servicios que son percibidos como seguros aportan mayor credibilidad a la información intercambiada, permiten alcanzar un mayor

nivel de cercanía, inmediatez y tangibilidad y generan entornos de confianza donde pueden participar clientes y empresas.

IV. DESARROLLO DEL TRABAJO

Para el desarrollo de la presente investigación se realizan una serie de pasos que se describen a continuación:

A. Análisis y selección de los casos de uso más relevantes.

En este paso definimos y caracterizamos una serie de servicios interactivos susceptibles de ser optimizados. De este modo, pretendemos identificar y seleccionar aquellos servicios que generan flujos de paquetes pequeños de larga duración, describiendo casos de uso en los que pueden emplearse dichos servicios dentro del ámbito comercial online. Por ejemplo, una plataforma de venta online o de soporte al cliente puede incluir un servicio de voz o videoconferencia con personal de la empresa.

B. Estudio de las expectativas del usuario final sobre los servicios interactivos seleccionados, profundizando en el valor y utilidad que espera obtener a partir de su empleo.

A través de este objetivo trataremos de determinar qué aspectos relativos a los servicios interactivos deben tenerse en cuenta desde la perspectiva del usuario para llevar a cabo su optimización y securización. De este modo, abordaremos las expectativas relacionadas con la utilidad, el valor generado y la seguridad, así como tipologías de productos en los que el empleo de estos servicios para fomentar la comunicación puede tener más éxito. Esto nos lleva a analizar el valor percibido por el usuario a partir del uso de los servicios interactivos, los requisitos de seguridad exigidos o la importancia de la confianza para emplear este tipo de servicio en un contexto comercial.

C. Avance en la definición de TCM.

Este objetivo pretende mejorar y documentar con mayor profundidad la propuesta del estándar TCM para la optimización del tráfico de paquetes pequeños. En la definición se incluye tanto la pila de protocolos a utilizar, como los parámetros de red que permiten mantener la calidad percibida en niveles adecuados. Se hará un especial énfasis en los esquemas y métodos para añadir seguridad, así como el estudio de los compromisos entre calidad, seguridad y eficiencia. Todo ello será estudiado en el contexto de los casos de uso definidos, incluyendo el análisis de su impacto en el usuario final, costes del uso y el análisis de la seguridad.

D. Desarrollo de una plataforma de pruebas.

Este objetivo persigue desarrollar una plataforma de pruebas que permita medir los parámetros de calidad de servicio (QoS) y de seguridad que se vean modificados por la optimización del tráfico de red en los casos de uso definidos en el proyecto. La plataforma debe permitir llevar a cabo pruebas con usuarios reales de los servicios interactivos, para poder así evaluar también la calidad percibida (QoE).

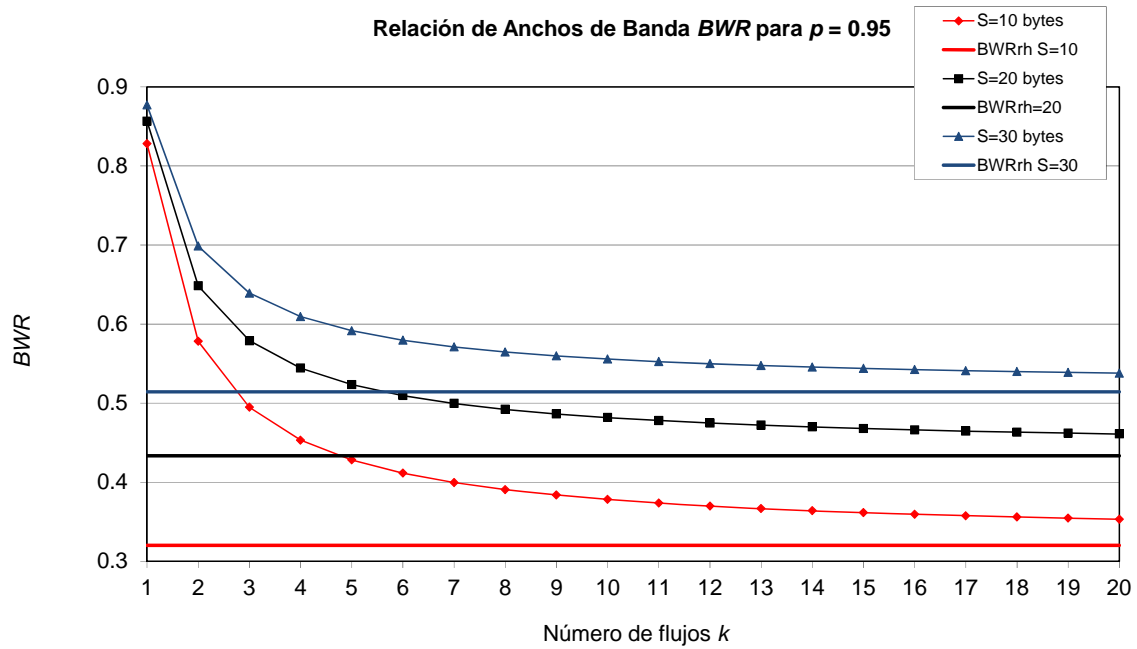


Fig. 3. Relación de anchos de banda BWR para una probabilidad de cabecera reducida de $\rho = 0,95$

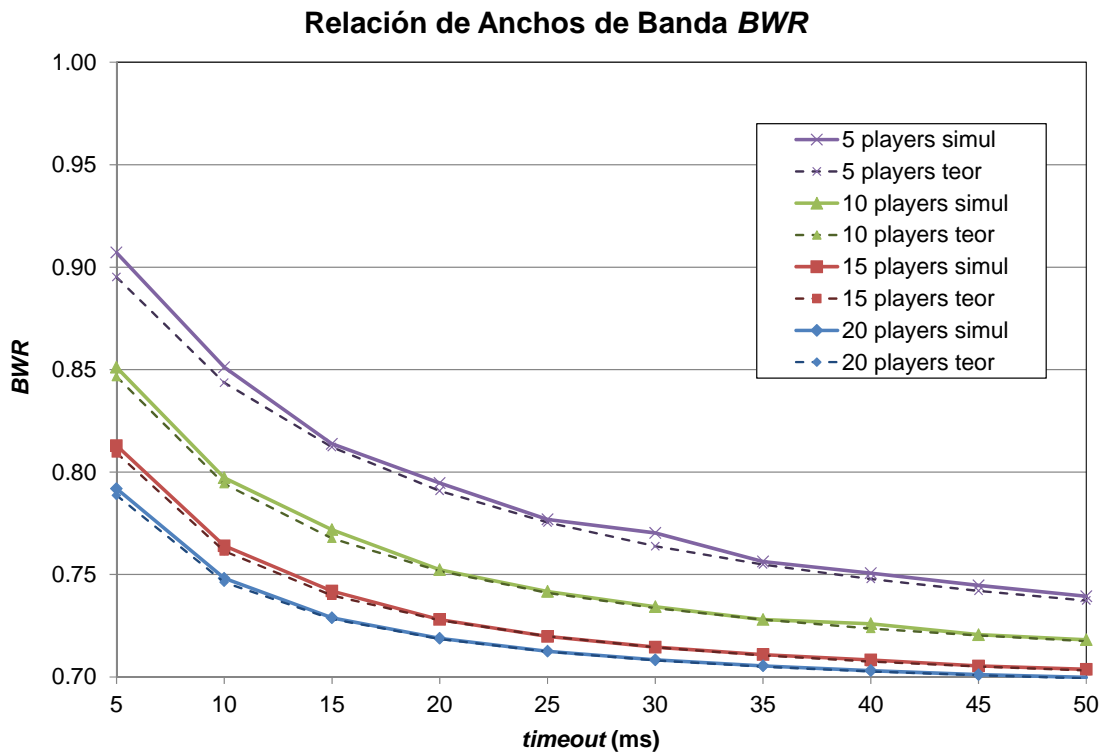


Fig. 4. Relación entre el ancho de banda nativo y el optimizado para el juego Counter Strike

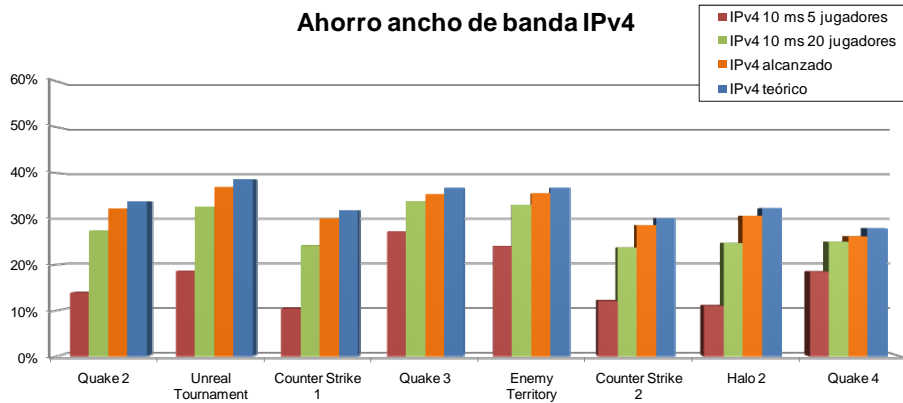


Fig. 5. Ahorro que se puede alcanzar en ocho juegos online

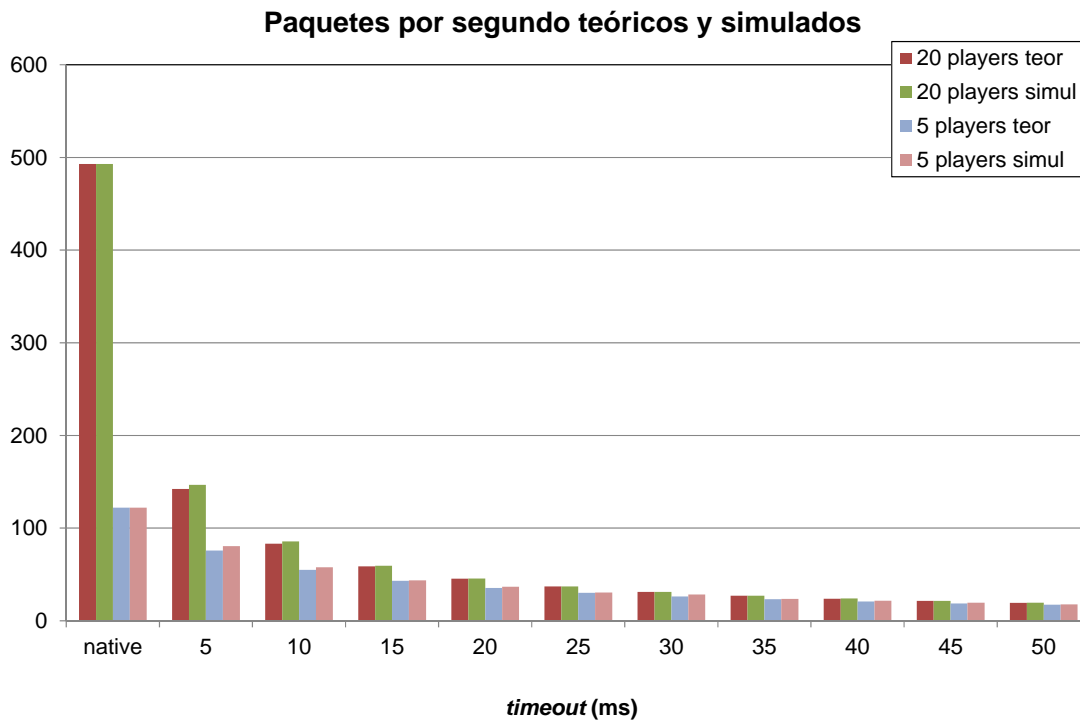


Fig. 6. Reducción en paquetes por segundo entre el tráfico nativo y el optimizado para el juego Counter Strike

E. Planificación de las pruebas y obtención de parámetros objetivos de la red.

Se llevarán a cabo las pruebas necesarias para evaluar la aplicabilidad de TCM. Además de medir los costes asociados en cuanto a inversión del nuevo equipamiento, capacidad de proceso y consumo energético en la red, se obtendrán resultados técnicos vinculados con la optimización del tráfico de los servicios interactivos y sus beneficios. Las pruebas realizadas en este punto abordarán aspectos como el porcentaje de reducción de ancho de banda al optimizar el tráfico, los límites de retardo tolerables para garantizar una calidad subjetiva adecuada, la influencia de los parámetros de la red (ancho de banda, retardo, pérdidas, jitter), la adecuación de distintos métodos de compresión de cabeceras según cada escenario (cableado o inalámbrico), etc.

A modo de ejemplo, las Fig. 3, 4, 5 y 6 obtenidas en trabajos previos [5],[20] muestran los ahorros de ancho de banda y paquetes por segundo que se pueden conseguir para dos servicios de tiempo real: VoIP y juegos online.

F. Análisis del impacto que los servicios interactivos optimizados ejercen en las percepciones y en el comportamiento del cliente de una plataforma comercial online.

Se analizarán aspectos derivados de las interacciones llevadas a cabo mediante los servicios desarrollados en el proyecto, como por ejemplo la calidad de la experiencia, el valor percibido, la satisfacción o la intención de participar activamente en la plataforma. De este modo, se evaluará si el cliente percibe más valor en la relación comercial al poder interactuar en tiempo real (mediante Voz sobre IP o

videoconferencia), frente al caso en que la comunicación se realiza, por ejemplo, en un foro.

G. Estudio del efecto que los servicios interactivos optimizados ejercen sobre la relación del cliente con la empresa.

En este objetivo analizaremos el efecto de la inclusión y empleo de este tipo de servicios en la gestión de marca de la empresa y en el valor total generado por la relación comercial. Igualmente, analizaremos el efecto de la pérdida de interactividad causada por el retardo en la optimización de la red.

V. CONCLUSIONES

A pesar del interés e importancia de los servicios multimedia interactivos, éstos apenas han sido incluidos en plataformas empresariales digitales, debido principalmente a que su desarrollo todavía adolece de ciertas limitaciones que les impide ofrecer la calidad deseada. Para cubrir estas limitaciones, surge la necesidad de aplicar un enfoque interdisciplinar, considerando, además del desarrollo técnico de este tipo de servicios, las expectativas y valoraciones del usuario final respecto a la optimización y securización planteada. Y en este sentido los objetivos que se proponen para abordar esta propuesta pueden resumirse en dos líneas principales:

- Profundizar en la propuesta de un estándar global de optimización de tráfico y su securización, manteniendo los requerimientos de calidad subjetiva y proporcionando el nivel de seguridad adecuado para el cliente.
- Evaluar la aplicabilidad de la optimización del tráfico y sus beneficios, midiendo el valor generado para sus usuarios finales, ya sean empresas o particulares. La posibilidad de establecer nuevos mecanismos de interrelación, a través de los cuales los clientes tengan la posibilidad de adquirir mayor información y de relacionarse en tiempo real, supone una fuente de valor fundamental para diferenciarse de la competencia y para generar rentas presentes y futuras.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por la Unión Europea a través del Proyecto Wi5 (H2020-644262), por la DGA a través de CeNITEQ y por la Universidad de Zaragoza a través del proyecto IHMOST (UZ2014-TEC-02).

REFERENCIAS

- [1] Huawei, "Smartphone Solutions White Paper", Issue 2, 2012.07.17., disponible online: http://www.huawei.com/ilink/en/download/HW_193034
- [2] J. Saldana, J. Fernandez-Navajas, J. Ruiz-Mas, J. Murillo, E. Viruete, J.I. Aznar. "Evaluating the Influence of Multiplexing Schemes and Buffer Implementation on Perceived VoIP Conversation Quality", *Computer Networks* (Elsevier), 56 (7), pp. 1893-1919, May 2012.
- [3] J. Park, B. Kim, Y.H. Choi "On the Efficacy PPPMux within Radio Access Network", *Lecture Notes in Computer Science*, 3262, pp. 521-530 Universal Multiservice Networks, Springer-Verlag 2004.
- [4] J. Saldana, J. Fernandez-Navajas, J. Ruiz-Mas, E. Viruete Navarro, L. Casadesus, "Influence of Online Games Traffic Multiplexing and Router Buffer on Subjective Quality", in *Proc. CCNC 2012*, pp. 482-486, Las Vegas. Jan 2012.
- [5] J. Saldana, L. Sequeira, J. Fernandez-Navajas, J. Ruiz-Mas, "Traffic Optimization for TCP-based Massive Multiplayer Online Games," *Proc. International Symposium on Performance Evaluation of Computer and Telecommunication Systems SPECTS 2012*, July 8-11, 2012, Genoa, Italy.
- [6] A. Abdul-Rahman. "The PGP Trust Model", *EDI-Forum*, April 1997.
- [7] Adi Shamir, *Identity-Based Cryptosystems and Signature Schemes*. *Advances in Cryptology: Proceedings of CRYPTO 84*, *Lecture Notes in Computer Science*, 7:47--53, 1984.
- [8] N.V. Wunderlich, F.V. Wangenheim, M.J. Bitner. "High Tech and High Touch A Framework for Understanding User Attitudes and Behaviors Related to Smart Interactive Services", *Journal of Service Research*, 16 (1), pp. 3-20, 2013.
- [9] A. Fano, A. Gershman. "The Future of Business Services in the Age of Ubiquitous Computing", *Communications of the ACM*, 45(12), pp. 83-87, 2002.
- [10] S.S. Shang, E.Y. Li, Y.L. Wu, O.C. Hou. "Understanding Web 2.0 service models: A knowledge-creating perspective", *Information & Management*, 48 (4), pp. 178-184, 2011.
- [11] C. Wang, P. Zhang. "The evolution of social commerce: an examination from the people, business, technology and information perspective", *Communication of the Association for Information Systems*, 31 (5), pp. 105-127, 2012.
- [12] J. Zhang, T. Daugherty. "Third-Person Effect and Social Networking: Implications for Online Marketing and Word-of-Mouth Communication", *American Journal of Business*, 24, pp. 53-63, 2009.
- [13] D. Godes, D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, P.W.J. Verlegh. "The firm's management of social interactions", *Marketing Letters*, 16 (3), pp 415-428, 2005.
- [14] B. Libai, R. Bolton, M. Bügel, K. de Ruyter, O. Götz, H. Risselada, A.T. Stephen. "Customer-to-customer interactions: Broadening the scope of word of mouth research", *Journal of Service Research*, 13 (3), pp. 267-282, 2010.
- [15] Harbor Research. "The 2013 Smart Systems Forecast", San Francisco, CA: Harbor Research, 2013.
- [16] M. Biehl, E. Prater, J.R. McIntyre. "Remote Repair, Diagnostics, and Maintenance", *Communications of the ACM*, 47 (11), pp. 100-106, 2004.
- [17] M.S. Yadavy, P. A. Pavlou. "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions", *Journal of Marketing*, 78 (1), pp. 20-40, 2014.
- [18] H.T. Keh, J. Pang, "Customer Reactions to Service Separation", *Journal of Marketing*, 74 (March), pp. 55-71, 2010.
- [19] J.H. Schumann, N.V. Wunderlich, F. Wangenheim. "Technology mediation in service delivery: A new typology and an agenda for managers and academics", *Technovation*, 32, pp. 133-143, 2012.
- [20] J. M. Saldaña Medina; et al. 2011. First Person Shooters: Can a Smarter Network Save Bandwidth without Annoying the Players?. *IEEE Communications Magazine*. 49 - 11, pp.190-198. ISSN 0163-6804.

Nuevas arquitecturas para la adaptación de servicios multimedia en redes móviles emergentes

Jose Oscar Fajardo, Ianire Taboada, Jose Ramón Otegi, Eva Ibarrola, Armando Ferro, Fidel Liberal

Departamento de Ingeniería de Comunicaciones

Universidad del País Vasco (UPV/EHU)

ETSI Bilbao, 48013 Bilbao, España.

{joseoscar.fajardo,ianire.taboada,joserra.otegi,eva.ibarrola,armando.ferro,fidel.liberal}@ehu.eus

Resumen- Este artículo trata sobre los esquemas de distribución de contenido multimedia a través de Internet, con especial interés en Internet móvil. Tanto la variabilidad de la calidad de los accesos radio como el incremento del tráfico multimedia plantean nuevos retos en relación a la gestión dinámica de los servicios sobre recursos de red compartidos. En primer lugar, el artículo ofrece una revisión del estado de la tecnología relacionada con el transporte multimedia extremo a extremo. A continuación, el artículo se centra en analizar la evolución de los mecanismos de adaptación dinámica, incluyendo desde esquemas extremo a extremo hasta la gestión dinámica de los recursos de red. Se analizan las posibles ventajas asociadas a nuevas arquitecturas como la computación en la nube móvil y las futuras redes 5G. Finalmente, el artículo identifica una serie de retos y desafíos en la gestión adaptativa de los servicios multimedia en el futuro.

Palabras Clave- servicios multimedia adaptativos, redes móviles emergentes, lógica de adaptación

I. INTRODUCCIÓN

En los últimos años se ha venido produciendo un incremento exponencial en lo relativo a la provisión de servicios y contenido multimedia a través de redes móviles de banda ancha. De acuerdo al informe “Ericsson mobility report” de noviembre de 2014, el 45% del tráfico en redes móviles multimedia está compuesto por contenidos multimedia, con un incremento esperado de 10X hasta el año 2020.

La distribución de contenidos multimedia a través de Internet conlleva una serie de requerimientos de calidad experimentada por el usuario del servicio –“Quality of Experience” (QoE)- que la red no es siempre capaz de proveer. Para hacer frente a esas carencias, los esquemas de transporte tradicionales extremo a extremo de tráfico multimedia han evolucionado de manera que los clientes son capaces de reaccionar a cambios en la conectividad manteniendo la continuidad de la reproducción. Desgraciadamente, estos mecanismos de adaptación basados en realizar cambios de nivel de calidad de codificación en los segmentos multimedia a solicitar conllevan un retardo de adaptación que puede resultar insuficiente en redes móviles. Este hecho suscita la propuesta de nuevas soluciones que introducen cierta lógica de adaptación en la red, con especial interés en el segmento de acceso radio.

Además, el tráfico de datos en Internet está experimentado un cambio significativo, de manera que cada vez mayor porcentaje del volumen total se corresponde a tráfico multimedia. Esto hace que la gestión de recursos de red compartidos resulte también cada vez más compleja.

A pesar de que las redes móviles de cuarta generación (4G) “Long Term Evolution” (LTE) y “LTE-Advanced” (lo que viene a denominarse comercialmente como 4G+) están ya siendo desplegadas a nivel internacional, ni el 3GPP ni otros organismos han definido un mecanismo universal de gestión de los recursos en sus estándares para redes de nueva generación. Con la transición de las redes móviles a redes orientadas a paquetes sobre recursos radio compartidos, la adecuada gestión de la QoE se hace cada vez más necesaria y a la vez se convierte en un desafío. Esta necesidad se verá aún más acentuada en las futuras redes móviles, teniendo en cuenta los principios planteados para las redes móviles de quinta generación (5G) como son la virtualización de recursos de red, la densificación de celdas o el “interworking” entre diferentes tecnologías de acceso radio (incluyendo redes celulares heterogéneas y redes WiFi).

Este artículo se centra en el análisis del estado actual de la gestión dinámica en el transporte de contenidos multimedia a través de Internet, así como de la evolución futura esperada.

La Sección II de este artículo recoge en el análisis de la evolución de los mecanismos de transporte de contenido multimedia en Internet, desde los protocolos de transporte en tiempo real hasta los mecanismos basados en protocolo HTTP. De igual manera, se realiza una revisión de los formatos de codificación actuales y su posible impacto en la transmisión a través de Internet. La Sección III describe la evolución histórica de los mecanismos de adaptación dinámica, tanto a nivel de servicio como a nivel de red. Así, se identifican diferentes arquitecturas de red emergentes que introducen mejoras en la provisión adaptativa de contenido multimedia a través de redes móviles. La Sección IV se centra en identificar un conjunto de retos de investigación asociados a estas nuevas arquitecturas y los desafíos que se plantean. Finalmente, la Sección V contiene las conclusiones al estudio.

II. EVOLUCIÓN DE LOS ESQUEMAS DE DISTRIBUCIÓN DE CONTENIDOS EN INTERNET Y REDES MÓVILES

Tradicionalmente, el tráfico multimedia se categoriza en contenido de tiempo real (por ejemplo, servicios conversacionales y “streaming” en tiempo real) y servicios de video bajo demanda –“Video on Demand” (VoD)-. El primero de los tipos de servicio suele ser provisto sobre transporte RTP/UDP debido a sus requerimientos de baja latencia. En el segundo caso, actualmente se prefiere el uso de transportes HTTP/TCP para descarga progresiva, en la

que los servidores ponen a disposición de los usuarios contenido pre-codificado y los clientes multimedia realizan una pre-bufferización del mismo para amortiguar posibles aumentos en los retardos de recepción del contenido [1]. Este esquema HTTP/TCP ofrece una serie de ventajas sobre el tráfico RTP/UDP relacionados con la actual arquitectura de Internet, como la posibilidad de uso de sistemas de caché de contenidos y un mejor comportamiento en escenarios de NAT y “firewalls” [2].

Más allá de la simple descarga progresiva HTTP, más recientemente se han desplegado esquemas de transmisión multimedia basados en dividir el contenido multimedia en segmentos de corta duración, de manera que cada segmento se puede descargar y reproducir de manera independiente. De esta manera, un mismo segmento se puede ofrecer en diferentes versiones de calidad y los clientes multimedia pueden elegir el formato de manera dinámica, pudiendo reaccionar rápida y automáticamente a variaciones en la calidad de la conexión. Esta capacidad resulta especialmente beneficiosa en entornos de acceso móvil, en los que la calidad del acceso puede variar bruscamente a lo largo del tiempo tanto por efectos radio como por el hecho de tener un canal compartido entre todos los usuarios móviles.

Este esquema ha sido estandarizado como “Dynamic Adaptive Streaming over HTTP” (DASH) en ISO/IEC 23009-1:2012 a partir de los trabajos en el grupo MPEG, y adoptado por el 3GPP como 3GP-DASH para servicios de “streaming” basados en conmutación de paquetes. DASH habilita el streaming adaptativo en bitrate, lo que permite a los clientes poder variar entre las diferentes calidades del contenido en base a la capacidad de recepción estimada de manera dinámica.

Actualmente, DASH suele usarse como vehículo para el transporte de contenidos multimedia en diferentes representaciones de calidad pero generalmente codificados en una sola capa mediante codecs como H.264/“Advanced Video Coding” (AVC) o H.265/“High Efficiency Video Coding” (HEVC).

En los últimos años, sin embargo, se ha visto incrementado el interés por el uso de codificación multi-capa en esquemas DASH. Este tipo de codecs, como el “Scalable Video Coding” (SVC), permiten codificar el contenido ofreciendo diferentes calidades de manera aditiva.

La Figura 1 ilustra las principales diferencias en el transporte DASH de contenidos multimedia con diferentes representaciones de calidad codificadas como representaciones independientes (Figura 1a) y con codificación multi-capa (Figura 1b).

En el ejemplo mostrado, cada segmento temporal de un flujo se ofrece en n versiones de calidad diferentes, asociados a diferentes bitrates de codificación. En el caso de AVC, para cada segmento se debe seleccionar la calidad deseada y solicitar el objeto HTTP correspondiente. Hasta que cada objeto completo no se recibe totalmente, no se puede continuar la reproducción. En el caso de SVC, se establece una capa base correspondiente a la calidad menor de las ofrecidas. Cada representación de calidad superior puede ser alcanzada solicitando el objeto HTTP incremental sobre el anterior.

De esta forma, un cliente puede decidir de manera más flexible la representación de calidad deseada ya que podría

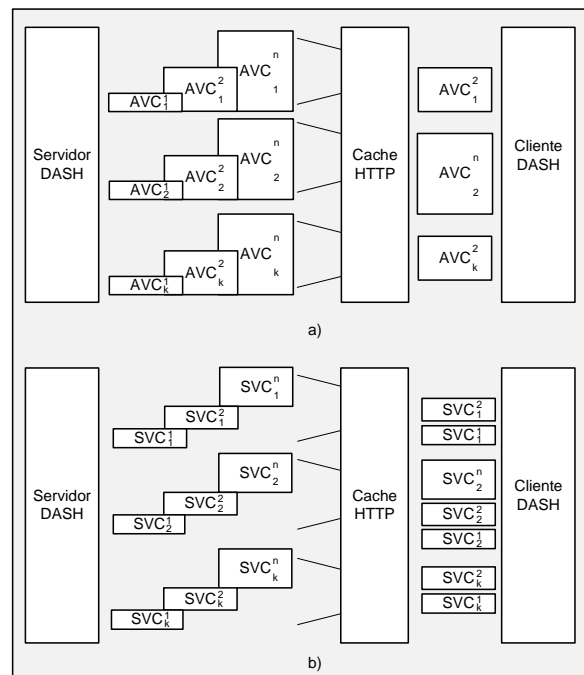


Fig. 1. Esquemas de transmisión DASH: a) AVC; b) SVC.

aumentar de calidad únicamente solicitando el objeto incremental sobre uno ya recibido en vez de necesitar el objeto completo para esa nueva calidad. Numerosos estudios [1, 2, 3, 4, 5] han analizado la mejora de rendimiento introducida por DASH-SVC gracias a esta flexibilidad.

III. EVOLUCIÓN DE LOS ESQUEMAS DE ADAPTACIÓN DE TRANSMISIONES MULTIMEDIA ORIENTADOS A INTERNET Y REDES MÓVILES

La Figura 2 representa la evolución histórica de diferentes arquitecturas de red orientadas a soportar la transmisión de contenidos multimedia habilitada con capacidades de adaptación dinámica en base a DASH. Como puede observarse, se han ido planteado diferentes puntos de adaptación a lo largo de la cadena de provisión de servicio de cara a optimizar la QoE de los usuarios finales con especial interés en entornos móviles.

A. Adaptaciones dirigidas por el cliente multimedia

En la Figura 2a) se representa el mecanismo de adaptación DASH tradicional, basado en que cada cliente selecciona en cada momento la representación de calidad que estima más oportuna para las condiciones de recepción dadas. Este esquema incluye un bucle de adaptación extremo a extremo entre el cliente y el servidor de contenidos. En este caso, el cliente va monitorizando la recepción de los objetos HTTP y, en función del número de objetos bufferizados, se solicita al servidor una determinada calidad para los siguientes segmentos multimedia.

Una mejora de esta arquitectura de transmisión es la colocación de cachés HTTP a lo largo de la cadena de provisión, organizados en torno al concepto de “Content Delivery Network” (CDN). Por un lado, esto mejora el rendimiento de las redes y del servicio. Por otro lado, al reducir el “Round Trip Time” (RTT) en el lazo de adaptación, mejora el tiempo de respuesta y el resultado de la

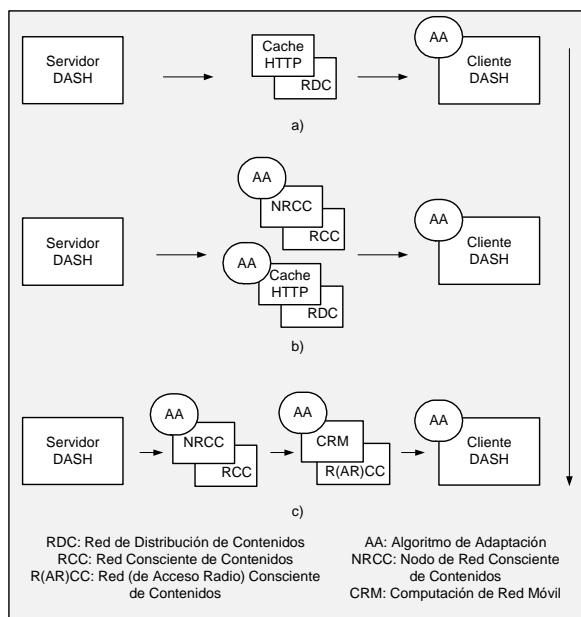


Fig. 2. Evolución de los esquemas de adaptación en distribución de contenidos multimedia: a) adaptaciones dirigidas por el cliente; b) adaptaciones asistidas por los nodos del núcleo de la red; c) adaptaciones asistidas por los nodos del núcleo de la red y de la red de acceso móvil.

transmisión adaptativa se traduce en contenidos con cambios de calidad más suavizados.

La literatura científica recoge diversos algoritmos de adaptación orientados a mejorar el rendimiento en diversos contextos de uso [1, 3, 5]. De igual forma, varios estudios se centran en el análisis de las ventajas e inconvenientes de utilizar AVC o SVC en transmisiones DASH [4, 5]. De manera resumida, se concluye que la utilización de DASH-AVC produce selecciones de calidad no óptimas en entornos de red con condiciones muy variables debido a que la elección de la representación debe hacerse a priori y con una determinada antelación en función del buffer de recepción utilizado. Una vez solicitado un segmento a determinada calidad, se debe esperar a la recepción total de dicho segmento para poder continuar con el proceso de reproducción. DASH-SVC ofrece tiempos de reacción más cortos ya que se puede cancelar la solicitud de un objeto correspondiente a una calidad alta sin interferir en la recepción de los objetos de calidad inferior. Además, DASH-SVC permite solicitar de manera dinámica un objeto incremental de calidad superior para un objeto ya recibido siempre que su recepción se estime para un tiempo anterior al instante de reproducción de dicho segmento multimedia.

Por tanto, se puede concluir que el uso de codificación multi-capa posibilita la adaptación dinámica de las transmisiones multimedia de manera mucho más flexible y rápida. Sin embargo, incluso en el caso multi-capa, cada cliente multimedia se comporta de manera individual y egoísta, de forma que el algoritmo de adaptación no es capaz de alcanzar un equilibrio estable en entornos con múltiples usuarios móviles con enlaces radio variables.

B. Adaptaciones asistidas por nodos de la red

La Figura 2b) ilustra la evolución de la arquitectura inicial de adaptación DASH extremo a extremo a una

arquitectura en la que los nodos de red son capaces de implementar cierta lógica de adaptación para los flujos multimedia que los atraviesan. Este tipo de mecanismos de adaptación DASH gestionados por la red han comenzado a ganar interés en la literatura científica [5, 6, 7, 8], de cara a optimizar la QoE en escenarios de transmisión multimedia a múltiples usuarios. El concepto básico subyacente radica en que estos elementos de red son capaces de modificar el flujo multimedia en función del rendimiento medido o estimado para el conjunto de los usuarios y no solo de forma individual, de manera que se puedan lanzar acciones de adaptación que permitan anticiparse a las acciones reactivas por parte del cliente multimedia. Estas acciones de adaptación gestionadas por la red, en función del tipo de codificación de los contenidos, pueden incluir desde transcodificación de segmentos multimedia a la selección automática y dinámica de la calidad que mejor se adapta al contexto actual de cada usuario.

Estas propuestas de optimización están generalmente orientadas a situaciones de codificación multi-capa y a su aplicación en los nodos del núcleo de la red. Desgraciadamente, una gran mayoría de estas propuestas siguen basándose en el uso de métricas de rendimiento de nivel de red o de servicio, y no permiten hacer uso de métricas de calidad más específicas de las redes móviles.

C. Adaptaciones asistidas por nodos de la red móvil

En la Figura 2c) se plantea la evolución lógica de este tipo de arquitecturas de adaptación gestionadas por la red con su inclusión en la futura Internet centrada en contenidos y en los principios de 5G. Los conceptos de “Network Function Virtualization” (NFV) y “Software Defined Networking” (SDN) hacen prever el despliegue de nodos de red flexibles de cara a ofrecer servicios más personalizados al contexto de usuario [9]. Un ejemplo reciente de cómo aplicar el paradigma SDN para asignar recursos de red de manera dinámica a diferentes clientes DASH puede encontrarse en [10], donde se plantea el uso de un gestor SDN para modificar los flujos de video con el objetivo de alcanzar un reparto justo de la QoE entre múltiples usuarios móviles.

Resulta de especial interés el despliegue de este tipo de nodos inteligentes en el extremo de la parte de red móvil, de cara a utilizar cierta información sobre la calidad de los enlaces radio en los algoritmos de adaptación. Si bien tradicionalmente ha sido realmente complicado introducir este tipo de lógica de servicio en el segmento de red móvil, actualmente se están planteando una serie de alternativas basadas en los principios de “Cloud Computing” que permitirían alcanzar estos objetivos.

En primer lugar, nos encontramos con la iniciativa “Mobile Edge Computing” (MEC), impulsada por un conjunto de grandes empresas del sector de las telecomunicaciones y planteada como grupo de especificación al ETSI [11]. Los principios básicos de esta arquitectura radican en el despliegue de un nodo servidor dentro de la “Radio Access Network” (RAN) que funcione como nexo de unión entre la parte móvil y de servicio. Así, el servidor MEC sería capaz de interactuar con los elementos radio del operador de red móvil y de proporcionar un espacio de procesamiento a servicios externos, de forma que estos servicios puedan ser desplegados de una manera totalmente personalizada y acomodados al contexto del usuario. Entre

otras, las capacidades avanzadas podrían incluir servicios de mínima latencia y máxima interactividad, servicios basados en la localización por red, servicios adaptados a la calidad radio instantánea, etc. En el caso de la arquitectura LTE, el servidor MEC estaría directamente integrado en el eNodoB.

Este servidor MEC, basado en principios de Cloud Computing, estaría orientado a ejecutar aplicaciones de terceros en un hardware de propósito general desplegado en la RAN. Dichas aplicaciones pueden incluir desde instancias de monitorización de los recursos radio asociados a los diferentes usuarios móviles, destinadas a alimentar la lógica de personalización de servicios desplegados externamente, hasta versiones ligeras de los propios servicios de distribución de contenido, capaces de gestionar y modificar el contenido hacia o desde los usuarios móviles a nivel RAN.

En segundo lugar, nos encontramos con la tendencia de evolución hacia redes 5G. Uno de los principios de esta evolución radica en la introducción de principios de virtualización hasta la propia RAN, en lo que viene a denominarse como "Cloud-RAN" (C-RAN) [12].

En el marco de esta nueva arquitectura se están planteando numerosas alternativas para la gestión optimizada de los recursos radio de manera personalizada y consciente de los contenidos y del contexto de los usuarios. Mientras que la alternativa MEC antes presentada se basa en gestionar los servicios cerca de los usuarios en función de su contexto, en la evolución 5G se plantea la gestión combinada de los servicios y de los recursos de red en base a los requerimientos y contexto de los usuarios móviles [13].

IV. RETOS DE INVESTIGACIÓN Y DESAFÍOS

En base al contexto tecnológico establecido previamente, se identifican una serie de retos de investigación que deberán ser afrontados en los próximos años de cara a introducir mecanismos de gestión de distribución de contenidos en redes móviles orientados a la QoE:

- Adaptación de servicios multimedia en redes móviles 4G+/5G, mediante el estudio de nuevos esquemas de codificación y formatos de transmisión multimedia, del impacto de las nuevas arquitecturas y tecnologías de redes móviles, y de la posibilidad de establecer puntos de adaptación de servicio a lo largo de la cadena de provisión. En este sentido, se deben definir nuevos algoritmos de adaptación que gobiernen los elementos de nivel de servicio en las redes modernas considerando las características de rendimiento variable en redes móviles y la satisfacción de los usuarios en el servicio.
- Gestión de recursos de red para la optimización de la QoE en redes 4G+/5G, mediante la aplicación de modelos analíticos que permitan el desarrollo de algoritmos de gestión de recursos de red orientados a la provisión óptima de contenidos multimedia. Este reto de investigación tiene un componente de corto plazo, relativo a la planificación avanzada de recursos radio 4G+, y otra más a medio plazo, orientada a la planificación avanzada de recursos de red virtualizados en 5G.

En las siguientes secciones se detallan los desafíos más importantes identificados para cada uno de estos retos.

A. Adaptación de servicios multimedia en redes móviles 4G+/5G

Este desafío está relacionado con en el estudio de nuevos algoritmos y mecanismos de adaptación de nivel de servicio que permita a las aplicaciones multimedia obtener un máximo rendimiento en términos de QoE en función del rendimiento de red experimentado en cada momento.

Los desafíos relacionados viene marcada por: (1) el despliegue comercial masivo de las redes LTE / LTE-A; 2) el despliegue masivo de nuevos mecanismos de distribución de contenidos, incluyendo soporte estandarizado en los navegadores web más utilizados; (3) la definición de nuevos formatos de codificación de "High-Definition" y "Ultra-High-Definition", etc. En el primer caso, de acuerdo a los últimos informes del GSA, actualmente hay 644 operadores invirtiendo en despliegues LTE para 181 países a lo largo de todo el mundo. Los despliegues más avanzados, incluyendo España, están ofreciendo ya LTE-Advanced en base a "Carrier Aggregation" con despliegues en diferentes bandas de frecuencia. Al mismo tiempo, los mecanismos de transmisión multimedia han evolucionando rápidamente hacia nuevos estándares DASH basados en HTTP, realizando la transición de anteriores formatos propietarios basados por ejemplo en Adobe Flash. Estos nuevos mecanismos, con la eventual irrupción de HTTP2.0 y protocolos alternativos de transporte, conllevan nuevos patrones de tráfico y nuevas relaciones entre el rendimiento de red y el rendimiento a nivel de aplicación. Finalmente, los formatos de presentación y codificación de los contenidos multimedia se encuentran igualmente en constante evolución, obteniéndose cada vez mejores relaciones de calidad a "bitrate" requerido.

Todo ello, junto a la explosión de los contenidos multimedia y su cada vez más elevado porcentaje en el tráfico de Internet móvil, hace pensar en un cambio importante en cuanto a los patrones de tráfico que se espera soportar en un futuro próximo en las redes móviles. Esta combinación de nuevos patrones de tráfico multimedia y el rendimiento aún incierto de las nuevas redes móviles LTE cuando estén plenamente instauradas y popularizado su uso, pone de relieve la necesidad de realizar nuevos estudios de rendimiento en términos de QoE y especialmente en cuanto a la eficacia de los mecanismos de adaptación de servicio.

Por tanto, para llegar al objetivo de definir algoritmos de adaptación a nivel de servicio que permitan optimizar la QoE de los usuarios en los nuevos contextos de servicios multimedia en función del rendimiento de red experimentado en cada momento, se plantean los siguientes desafíos:

- Modelado de patrones de tráfico. El modelado de los patrones de tráfico asociados a los nuevos mecanismos de transmisión y formatos de codificación es crítico de cara a estudiar correctamente el impacto de los segmentos radio en las aplicaciones (ej., [14] en redes 3G).
- Modelado de canal 4G+ / 5G. Se plantea la necesidad de generar modelos de rendimiento en entornos de despliegues reales sometidos a movilidad. Se deben estudiar los diferentes efectos debidos a la transmisión de las señales radio, a los procesos de scheduling en los eNodoB, a los mecanismos HARQ, etc. Se debe primar el estudio sobre despliegues reales comerciales, que son en definitiva los que

marcan la QoE de los usuarios [15]. En el caso de no disponer de despliegues reales, como para las futuras redes 5G, se pueden emplear modelos que incluyan los nuevos esquemas radio de cara a estimar el rendimiento de las conexiones móviles.

- Impacto en las aplicaciones multimedia. Se plantea analizar el impacto combinado a nivel de aplicación y a nivel de usuario en términos de QoE de los perfiles de tráfico multimedia en diferentes condiciones de red. Se hace necesario realizar un mapeo de los parámetros de rendimiento a nivel radio, IP y los parámetros de rendimiento a nivel de aplicación en función de diferentes parámetros de servicio [16].
- Diseño de los mecanismos de adaptación. Los algoritmos de control que permitan adaptar los parámetros de servicio para optimizar la QoE deben tener en cuenta tanto el contexto de servicio como los parámetros radio específicos en entornos 4G+ / 5G. Resulta de especial interés la incorporación de modelos de inteligencia artificial, entrenados con las condiciones de servicio estudiadas considerando los ciclos de histéresis asociados a los niveles de calidad subjetiva promedio en entornos variables [17].

B. Gestión de recursos de red para la optimización de la QoE en redes 4G+/5G

Los desafíos planteados en el punto anterior tratan sobre algoritmos de optimización de nivel de servicio que se ejecutan individualmente en cada cliente multimedia, de cara a maximizar la QoE de cada usuario en función de sus capacidades instantáneas. En entornos con múltiples usuarios con características de rendimiento radio muy variable, este tipo de esquemas ofrecen soluciones sub-óptimas ya que las acciones de unos usuarios influyen en el resto de usuarios debido al uso compartido de recursos radio.

En este punto se identifican una serie de desafíos relacionados con la gestión de recursos de red, con especial foco en redes móviles, que permitan maximizar la QoE del conjunto de los usuarios móviles de una manera coordinada.

En el contexto tecnológico analizado anteriormente, se destacaban dos arquitecturas como principales evoluciones a la actual Internet móvil. Por un lado, la propuesta de Mobile Edge Computing permite situar la distribución personalizada de contenidos multimedia lo más cerca posible de los usuarios móviles, utilizando recursos compartidos hardware en la RAN. Por otro lado, los principios establecidos para 5G incluyen la integración de un conjunto de evoluciones tecnológicas entre los que se pueden citar: el uso de recursos de red virtualizados y compartidos (desde el “backhaul” móvil hasta la C-RAN), el uso de nuevos esquemas de modulación con tasas de transferencia y latencias mejoradas, tecnologías radio de MIMO masivo, el uso de accesos radio multi-tecnología (incluyendo diferentes tecnologías de red móvil de banda ancha, conectividad WiFi, redes WiFi comunitarias, etc.)... Si bien se están comenzando a estandarizar las arquitecturas que permitirán la coordinación de todos estos aspectos heterogéneos (por ej. ETSI MANO [18]), hay mucho trabajo que realizar en lo relativo a la adecuada orquestación de recursos para la provisión de servicios extremo a extremo con requerimientos de QoE.

Dentro de esta amalgama de conceptos tecnológicos, se destacan dos retos de investigación de cara a progresar en la

organización óptima de recursos de red con el objetivo de maximizar la QoE en servicios multimedia móviles:

- Políticas de scheduling en un nodo para la provisión óptima de servicios multimedia

Este desafío contempla el uso de un tipo de algoritmos de optimización que se basan en maximizar el beneficio común en función de los requerimientos de QoE y de las características radio de cada usuario en cada momento [19, 20]. Por tanto, la implementación de estos algoritmos requiere que el nodo en cuestión monitorice tanto parámetros de nivel de servicio como de rendimiento radio.

Se plantean diferentes casos de uso en función de las diferentes arquitecturas de red móvil planteadas: En primer lugar, se plantea el desafío de incluir lógica inteligente para gobernar el “scheduling” de un eNodoB en la arquitectura LTE actual, teniendo en cuenta que pueden servirse usuarios de diferente tipo (por ej., UEs con soporte LTE Categoría 4 y LTE-A Categoría 6) y con diferentes capacidades y requerimientos. En este caso, hay que tener muy presentes las limitaciones impuestas por el hardware de los eNodoB. En segundo lugar, se plantea el reto de incluir lógica inteligente para gobernar el “scheduling” de un eNodoB virtualizado en la arquitectura C-RAN, que puede permitir una mayor flexibilidad y capacidad de procesamiento. En tercer y último lugar, se plantea el reto de incluir lógica inteligente para gobernar el scheduling de una instancia de servicio multimedia en la arquitectura MEC, teniendo en cuenta el conocimiento parcial e inexacto del canal radio de los usuarios móviles.

- Algoritmos de optimización de redes móviles para la provisión óptima de servicios multimedia

Este desafío extiende la problemática de gestión de recursos de red en entornos multi-usuario a escenarios multi-nodo. Por tanto, este tipo de algoritmos están orientados a la planificación y gestión dinámica de la cadena de provisión de servicios de una manera coordinada [10, 13, 21, 22].

A continuación, se plantean una serie de escenarios como referencia de problemas que deben ser resueltos de cara a una gestión adecuada de las futuras redes móviles. En primer lugar, la arquitectura ETSI MANO define un elemento orquestador de recursos virtualizados. En el contexto de redes 5G, se debe avanzar en la lógica inteligente que permita la provisión de servicios multimedia con requerimientos de QoE, con especial interés en la parte de coordinación de recursos radio en la C-RAN. Como caso particular, se plantea la gestión de servicios bajo accesos radio multi-tecnología. En este campo se pueden incluir desde la coordinación de recursos en arquitecturas LTE-A con agregación de portadoras de diferentes celdas, hasta la orquestación en accesos con diferentes tecnologías radio (por ejemplo, LTE y WiFi). Finalmente, se plantea el concepto de despliegues densificados como otro aspecto clave de cara a mejorar la QoE en las futuras redes 5G. Por un lado, el uso de celdas de pequeña cobertura o “Small Cells” permite ofrecer servicios de conectividad radio mejorados frente a celdas macroscópicas. Sin embargo, esta arquitectura de múltiples puntos de acceso plantea nuevos desafíos en la gestión coordinada de recursos, y especialmente en la continuidad de servicio para usuarios en movilidad. Este desafío incluye conceptos como la anticipación a la actividad del usuario móvil y la pre-asignación de recursos radio virtualizados.

V. CONCLUSIONES

Este artículo ofrece en primer lugar una revisión sobre el estado actual de las tecnologías asociadas a la distribución de contenidos multimedia a través de Internet, con especial interés en el acceso mediante redes móviles. Se analizan las ventajas de utilizar protocolos de transmisión adaptativos, que permiten a los clientes multimedia ejecutar una lógica orientada a evitar cortes en la reproducción mediante un control dinámico de la calidad solicitada. Igualmente, se revisan las ventajas de incluir formatos de codificación multi-capas en este tipo de transmisiones, especialmente relevantes en escenarios de video bajo demanda.

Posteriormente, se analizan las ventajas e inconvenientes de diferentes esquemas de adaptación que permiten reaccionar a situaciones de degradación del rendimiento en el acceso a Internet. Por un lado, las adaptaciones implementadas en los clientes suelen conllevar unos tiempos de adaptación que no siempre permiten reaccionar adecuadamente a cambios bruscos de calidad como los que pueden encontrarse en redes móviles. Si bien la inclusión de formatos de codificación multi-capas puede aportar mayor flexibilidad, la falta de un control coordinado hace que las adaptaciones de cliente no ofrezcan resultados de calidad óptimos en entornos multi-usuario.

Es por ello que en los últimos tiempos se han planteado una serie de propuestas de gestión con soporte de la red. Por un lado, se plantea el uso de nodos de red conscientes del contenido y del contexto de los diferentes usuarios, de manera que puedan modificar el flujo multimedia con el objetivo de maximizar la calidad del conjunto de los usuarios. Por otro lado, se plantea el problema de la adecuada asignación de los recursos de red disponibles a los diferentes flujos multimedia, igualmente con el objetivo de maximizar la calidad del conjunto de los usuarios.

Aunque este tipo de esquemas mixtos de adaptación en cliente y en red proporcionan ciertas mejoras, hasta hace poco no se ha planteado la posibilidad de incluir este tipo de elementos de gestión en la red de acceso móvil. Actualmente, se está estudiando la aplicación de una serie de nuevas arquitecturas emergentes, como la computación compartida en la red móvil o el uso de redes de acceso móvil con recursos virtualizados. Las ventajas de acercar la lógica de adaptación multimedia al segmento de red móvil son significativas. Por un lado, se pueden acomodar los flujos multimedia al rendimiento de red instantáneo de cada usuario. Por otro lado, los tiempos de reacción asociados se acortan hasta los propios límites inevitables de la parte radio.

Como puede observarse, nos encontramos ante un escenario realmente interesante, pero que a la vez introduce una serie de desafíos que se deben afrontar en los próximos años. En este artículo, se tratan de identificar los retos más importantes relacionados con la lógica de adaptación dinámica multimedia tanto a nivel de servicio como a nivel de red.

AGRADECIMIENTOS

Este trabajo ha sido en parte financiado por el Ministerio de Economía y Competitividad (MINECO) del Gobierno de España bajo el proyecto TEC2013-46766-R: QoEverage - Mecanismos basados en la QoE para la optimización de redes y servicios de nueva generación.

REFERENCIAS

- [1] C. Sieber, T. Hosfeld, T. Zinner, P. Tran-Gia, and C. Timmerer, "Implementation and User-centric Comparison of a Novel Adaptation Logic for DASH with SVC", En Actas de IFIP/IEEE International Symposium on Integrated Network Management, 2013, pp. 1318–1323.
- [2] Y. Sánchez de la Fuente et al., "iDASH: Improved Dynamic Adaptive Streaming over HTTP using Scalable Video Coding", En Actas de Second annual ACM conference on Multimedia systems, 2011, pp. 257–264.
- [3] C. Muller, D. Renzi, S. Lederer, S. Battista, and C. Timmerer, "Using Scalable Video Coding for Dynamic Adaptive Streaming over HTTP in mobile environments", En Actas de IEEE 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 2208–2212.
- [4] J. Famaey, S. Latré, N. Bouten, W. Van de Meerssche, B. De Vleeschouwer, W. Van Leekwijck, and F. De Turck, "On the merits of SVC-based HTTP adaptive streaming", En Actas de IFIP/IEEE International Symposium on Integrated Network Management, 2013, pp. 419–426.
- [5] R. Huysegems, B. De Vleeschouwer, T. Wu, and W. Van Leekwijck, "SVC-Based HTTP Adaptive Streaming", Bell Labs Technical Journal, vol. 16, n. 4, pp. 25–41, 2012.
- [6] N. et al Bouten, "Improved delivery of live SVC-based HTTP adaptive streaming content", En Actas de IEEE Network Operations and Management Symposium (NOMS), 2014, pp. 1–2.
- [7] M. Grafl et al., "Scalable Media Coding Enabling Content-Aware Networking", IEEE MultiMedia, vol. 20, n. 2, pp. 30–41, 2013.
- [8] S. Lederer, C. Mueller, C. Timmerer, H. Hellwagner, "Adaptive Multimedia Streaming in Information-Centric Networks", IEEE Network, November/December 2014.
- [9] D. Soldani, and A. Manzalini, "Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society", IEEE Vehicular Technology Magazine, vol. 10, n. 1, pp. 32–42, 2015.
- [10] P. Georgopoulos, Y. Elkhatib, M. Broadbent, M. Mu, and N. Race, "Towards Network-wide QoE Fairness Using OpenFlow-assisted Adaptive Video Streaming", En Actas de ACM SIGCOMM workshop on Future human-centric multimedia networking (FhMN '13), 2013, pp. 15–20.
- [11] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud Radio Access Network (C-RAN): A Primer", IEEE Network, vol. 29, n. 1, pp. 35–41, 2015.
- [12] M. Patel et al., "Mobile-Edge Computing Introductory Technical White Paper", Sept. 2014.
- [13] C.-F. Lai, R.-H. Hwang, H.-C. Chao, M. M. Hassan, and A. Alamri, "A Buffer-Aware HTTP Live Streaming Approach for SDN-Enabled 5G Wireless Networks", IEEE Network, vol. 29, n. 1, pp. 49–55, 2015.
- [14] J.J. Ramos-Muñoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, J.M. Lopez-Soler, "Characteristics of mobile youtube traffic", IEEE Wireless Communications, vol. 21, n. 1, pp. 18–25, 2014.
- [15] J.O. Fajardo, I. Taboada, F. Liberal, "Analysis of CQI traces from LTE MIMO deployments and impact on classical schedulers", En Actas de 13th International Conference on Wired & Wireless Internet Communications, 2015.
- [16] N. Becker, A. Rizk, M., Fidler, "A measurement study on the application-level performance of LTE", En Actas de IFIP Networking Conference, 2014.
- [17] A. Khan, L. Sun, E. Iteador, J.O. Fajardo, F. Liberal, H. Koumaras, "Video Quality Prediction Models Based on Video Content Dynamics for H.264 Video over UMTS Networks", International Journal of Digital Multimedia Broadcasting, vol. 2010, 2010.
- [18] M. Ersue, "ETSI NFV Management and Orchestration – An Overview", 2014.
- [19] I. Taboada, F. Liberal, J.O. Fajardo, U. Ayesta, "QoE-aware optimization of multimedia flow scheduling", Computer Communications, vol. 36, n. 15, pp. 1629–1638, 2013.
- [20] K. Ivesic, L. Skorin-Kapov, M. Matijasevic, "Cross-layer QoE-driven admission control and resource allocation for adaptive multimedia services in LTE", Journal of Network and Computer Applications, vol. 46, pp. 336–351, 2014.
- [21] J. Seppänen, M. Varela, A. Sgora, "An autonomous QoE-driven network management framework", J. Vis. Commun. Image R., vol. 25, pp. 565–577, 2014.
- [22] S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, D. Hock, M. Jarschel, M. Hoffmann, "Heuristic Approaches to the Controller Placement Problem in Large Scale SDN Networks", IEEE Transactions on Network and Service Management, vol. 12, n. 1, pp. 4–17, 2015.

Wireless and Multimedia Networking (WiMuNet) Laboratory

Jorge Navarro-Ortiz, Pablo Ameigeiras, Juan J. Ramos-Muñoz,
Jonathan Prados-Garzon, Juan M. Lopez-Soler
Dpt. of Signal Theory, Telematics, and Communications,
ETSI Informática y de Telecomunicación
Universidad de Granada (Spain)
{jorgenavarro, pameigeiras, jjramos, jpg, juanma}@ugr.es

Abstract- This paper reviews the research group *Wireless and Multimedia Networking Lab*, which is formed by faculty and researchers of the Department of Signal Theory, Telematics and Communications of the University of Granada. Its research lines include *broadband wireless networks, video streaming and multimedia traffic analysis, Software Defined Networking, Quality of Experience, Wireless Local Area Networks and Data Distribution Service*. Our latest research projects are focused in 5G mobile communication architectures based on *Software Defined Networking and Quality of Experience for multimedia services*. Teaching activities of the group members as well as a short bios are also included.

Palabras Clave- WiMuNet, 5G, wireless, multimedia, SDN, DDS.

I. INTRODUCTION

This paper presents the Wireless and Multimedia Networking (WiMuNet) Lab [1] of the University of Granada. It summarizes our activities and achievements in research and training.

WiMuNet is a research team recognized and partially financed by the Andalusian Scientific Information System [2], with identification label TIC-235. It was born in year 2012 as a split from the former Research Group in Signals, Telematics and Communications (TIC-123), founded in 1990. Headed by Prof. Lopez-Soler, WiMuNet members are currently faculty and Ph.D. students from the Higher Technical School of Computer Science and Telecommunications Engineering of the University of Granada, in the area of Telematics Engineering at the Dpt. of Signal Theory, Telematics and Communications. Some of WiMuNet members accumulate past professional experience working in leading companies as Nokia Networks, Optimi and Siemens.

To provide an overall overview of WiMuNet, the rest of the paper is organized as follows. Section II presents the main research lines of the group. Section III summarizes the latest research projects and contracts, whereas Section IV details our teaching activities. Sections V-VIII includes the awards, Ph.D. thesis, patents and short bios of the WiMuNet members. Section IX concludes the paper.

II. RESEARCH LINES

This section presents the main research lines of the WiMuNet Lab.

A. Broadband wireless networks

The area of broadband wireless communication networks is devoted to research and design of 4G (namely, LTE and LTE-Advanced) and 5G mobile networks. Its core objective includes the research of coverage improvements and capacity enhancement solutions that are essential for broadband wireless networks in order to deliver cost-efficient wireless services. Additionally, we investigate the satisfaction of future mobile communication networks requirements through the adoption of an SDN-based design of a hierarchical architecture for the 5G packet core.

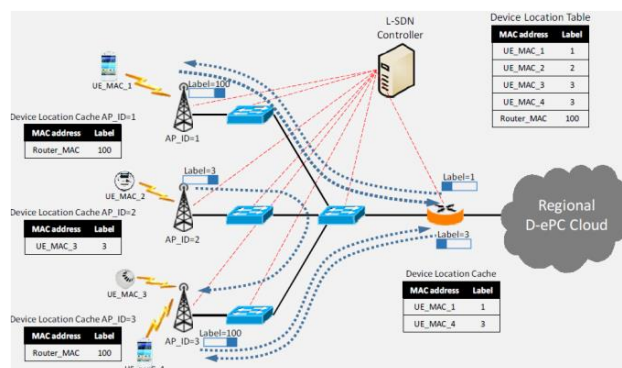


Fig. 1. 5G Access Cloud Architecture.

5G packet core design- The future data traffic load estimations forecast x1000 bandwidth increase in wireless networks by 2020. It will be caused by enriched and augmented new applications in smartphones, as well as massive deployments involving huge number of smart devices using Machine Type Communications (MTC). This disruptive demand is considered a major challenge for future mobile networks. Analysis of this traffic increase reveals new types of applications with extremely challenging requirements, such as vehicular communications or medical applications. The challenges that future 5G systems must address can be summarized as: 1) system capacity and data rate x1000, 2) massive number of connections due to MTC, 3) “zero-perceived” latency, 4) CAPEX and OPEX cost reductions, and 5) quality of experience optimization depending on the type of application. In this context, WiMunet has proposed (Fig. 1) [3][4] the design of an

Access Cloud architecture based on the SDN paradigm with the aim of fulfilling two main link-level challenges: mobility support and network scalability.

Small cells deployment in TV White Spaces– The operation of small cells in TV white spaces represents a coexistence challenge due to their unplanned deployment, their heterogeneous transmission technologies, and the scarcity of TVWS channels in crowded cities. We have designed [5] a centralized game-theoretical solution based on the IEEE 802.19.1 system that maximizes the average network capacity and simultaneously minimize the number of reconfigurations in neighbor cells when a new cell is turned on, based on the *neighborhood* concept (Fig. 2).

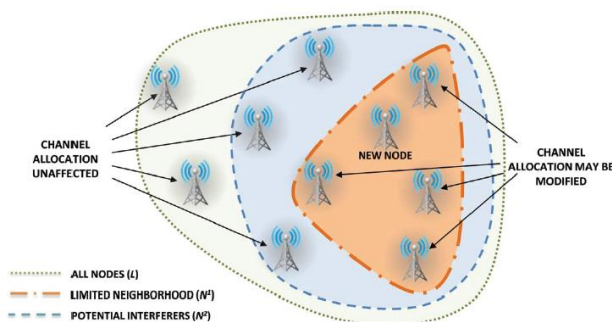


Fig. 2. Channel allocation on a limited neighborhood.

Traffic Models in OFDMA– Our research also studied the impact on the design of scheduling algorithms for Orthogonal Frequency Division Multiple Access (OFDMA) systems of two traffic models: the full buffer and the finite buffer traffic models. In [6] the analysis concentrates on utility-based scheduling with an α -fair utility function for Non-Real Time (NRT) services. Our results show (Fig. 3) that a gradient scheduling algorithm is able to maximize the aggregate utility over all the users when the less realistic full buffer model is adopted; but not when the finite buffer model is applied.

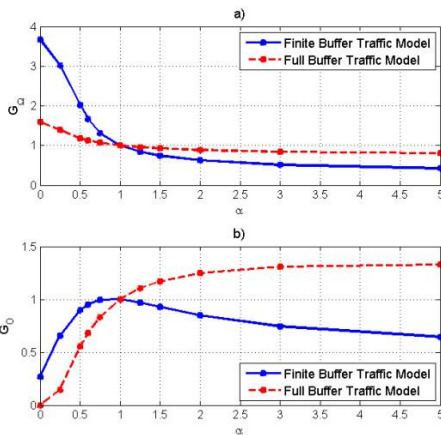


Fig. 3. Finite and full buffer traffic models comparison. (a) Average user throughput gain over Proportional Fair G_{Ω} . (b) User throughput at 5% outage gain over Proportional Fair G_{Ω} for different α values.

Packet scheduling for OFDMA-based networks– In this case, our research focuses on the performance evaluation (in

terms of cell throughput and coverage) and design of packet scheduling algorithms (Fig. 4) in OFDMA-based systems by means of network-level Montecarlo simulations. Evaluation is carried out under different degrees of fairness, traffic models, QoS requirements, or propagation conditions [7][8].

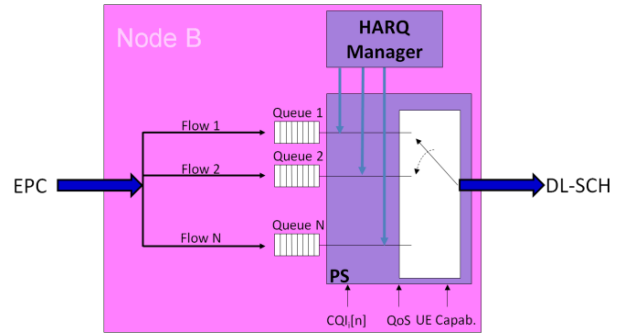


Fig. 4. Packet scheduling for OFDMA-based networks.

B. Video streaming and multimedia traffic analysis

Although we have several contributions related to multimedia services, such as e.g. automatic protocol selection [9], low-latency packet inter-leaver for VoIP using active networks [10-11], or service enhancements using active queue management [12], our most remarkable contributions in this field are related to the analysis of traffic patterns from YouTube servers (Fig. 5). In particular, we have provided traffic models when the video is downloaded from a computer [13] or to a mobile [14] (iOS and Android) device.

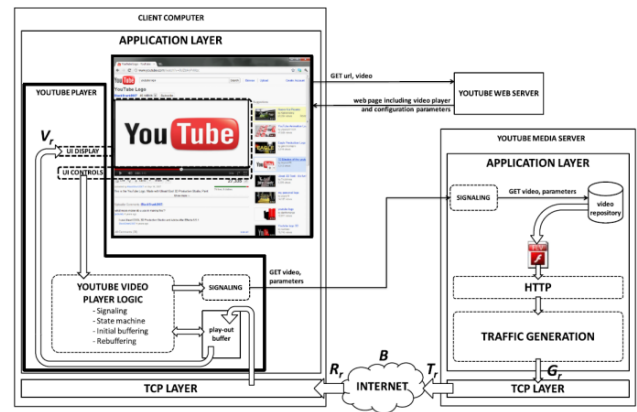


Fig. 5. Overview of the operation of the YouTube service.

In addition, we have proposed a simple model for predicting the number of rebuffering events and their duration in progressive downloads from YouTube. These metrics are necessary to predict the quality perceived by YouTube users. The proposed model can be easily implemented in simulation tools. In [15] we present an example of its use in a Long-Term Evolution simulator.

We also studied the performance of the YouTube service over 3G Long Term Evolution (LTE). We considered a typical configuration of an LTE network for TCP traffic and the traffic generation model for YouTube ‘Flash’ videos downloaded onto a personal computer (PC). The results obtained show that: the number of pauses experimented by users during video download are heavily influenced by the

cell load, but the same is not true for pause duration; most of the packet losses occur during initial burst due to the TCP adaptation to the BDP of the link, and, unlike packet losses during throttling phase, these are not depend on radio link quality; and in most cases the user do not use the maximum data rate achievable in the LTE interface [16].

C. Software Defined Networking

The Information and Communication Technologies (ICT) industry is witnessing a radical paradigm shift with the commoditization of hardware resources and the adoption of cloud computing. In the computer networking field, Software Defined Networking (SDN) and Network Function Virtualization (NFV) are achieving significant success. In particular, both SDN and NFV are expected to play a relevant role in 5G networks [3].

Design and evaluation of OpenFlow services for the provision of QoE– Although SDN is been embraced to provide great flexibility at low expenses, at this moment it has not standardized QoS support. However, SDN allows the development of agile and dynamic services which could benefit the overall QoE assessment of the multimedia services in a network.

To validate this thesis, we are designing and implementing a service which creates QoS-aware routes as part of the OpenDayLight controller. To that end, the controller uses the payload type of a RTP (Real-Time Protocol [18]) flow, calculates an optimal route based on de delay, bandwidth and packet loss rate requirements of the type of content (video streaming or voice over IP). When the network topology changes, the routes for the active flows are updated.

Additionally, we are designing and evaluating several schemes for providing QoS and quality of experience (QoE) to several popular multimedia services in SDN networks, while maintaining the compatibility with legacy (nowadays) networks. To this end, we will define network elements (that will be referred to as middleboxes hereafter) which can be integrated into present and SDN networks.

D. Quality of Experience

Emergent multimedia applications and services, such as VoIP, audio and video streaming, etc., have to face the impairments that IP networks impose (jitter, unbound end-to-end delays, consecutive packet losses, etc.). These impairments can degrade rapidly the quality of the aforementioned services, even making them unusable. Therefore, mitigating the impact of the network impediments and the measurement of such impact, is a must in IP networks.

From the point of view of the network service provider, one of the most important measures of the service performance is the final user satisfaction level. The modeling of the Quality of user Experience (QoE) implies the mapping of objective metrics (such as network impairments, signal degradation or service performance) into an opinion score. These models, tailored for each class of service, may serve to assess the QoE of a service performance in an automatic way. Moreover, using the QoE models as utility functions, we can devise automatic adaptive algorithms to control protocols

selection, resource management or reparation schemes operation.

Utility function for web browsing– After conducting a number of web browsing surveys, we have derived a utility function (Fig. 6) which maps the whole page download time to the user satisfaction (*mean opinion score*) [19].

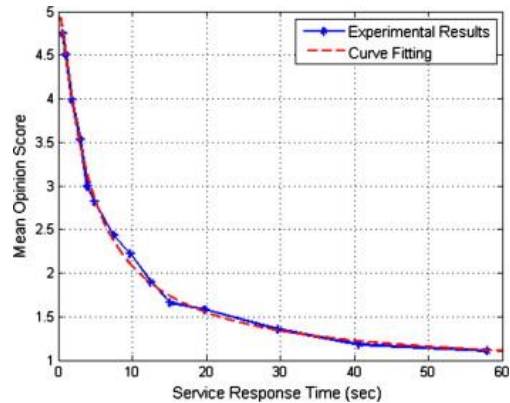


Fig. 6. MOS Vs download time for web browsing.

Utility function for YouTube– Based on surveys, we have derived a utility function which maps the number and duration of playback interruptions (buffer stalls) to the user satisfaction.

QoE-based packet scheduling for LTE and WLAN networks– Based on derived utility functions (and others from the literature), we have designed several mechanisms and algorithms to optimize the end-users perception, especially for LTE [7][8][19] and IEEE 802.11 networks [20][21].

Lightweight Autonomous resilient Networks (LAIN)– LAIN is a proof of concept that a network can select the most promising protocol automatically given dynamic network conditions, by means of probing a set of protocols with provides the same service, and evaluating their performance for each instant. To get this architecture to work, the proper utility functions should be provided to guide the protocol selection [9].

E. Wireless Local Area Networks

We have active research in the field of Wireless Local Area Networks, focusing on quality of service (QoS) and experience (QoE) mechanisms, both in low layers (especially in the MAC layer) and high layers (optimization of TCP/IP protocols and crosslayer procedures) .

QoE-based radio resource management for the IEEE 802.11e HCCA mechanism– The standard allows the developer to select the packet scheduler that best fit his purposes as long as some service requirements are fulfilled. Based on the Monolithic Shaper-Scheduler, we have proposed a method that modifies the resources granted for each traffic class (Fig. 7) according to operator-defined policies, e.g. to maintain certain required levels of Quality of

Experience, to achieve user fairness or to optimize operator’s revenue [20][21].

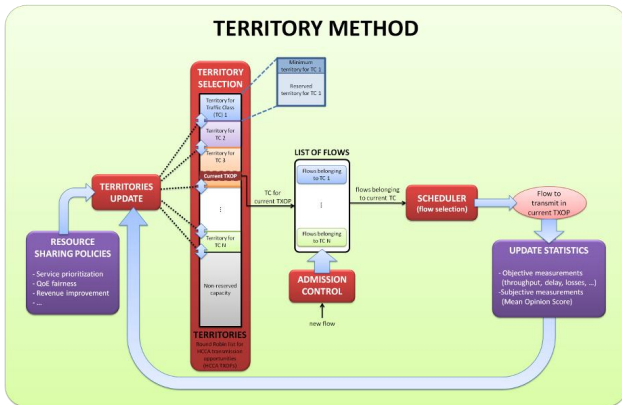


Fig. 7. Territory Method diagram.

TCP enhancement for WLAN– Based on the HCCA mechanism, we have developed [22] a TCP flavor, *Split TCP and UDP (STU)* (Fig. 8), which avoids unfairness between uplink and downlink traffic while improving the end-user performance.

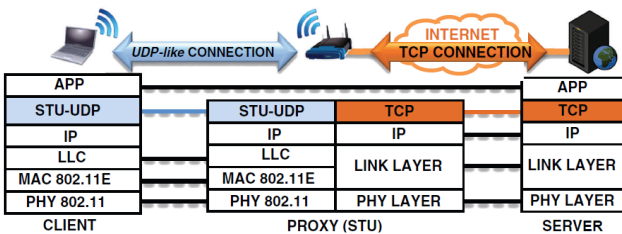


Fig. 8. STU in a wired-cum-wireless network.

F. Data Distribution Service

The Data Distribution Service (DDS) is an OMG standard for data-centric publish-subscribe (DCPS) communication systems. Our research in this field is focused in improving some aspects of the DDS middleware such as discovery, scalability, cloud monitoring, and the use of this technology for real-time multimedia streaming. Our laboratory is partner of Real-Time Innovations Inc., which is “the most influential industrial Internet of Things company” [23]. We are also members of the RTI Research Community.

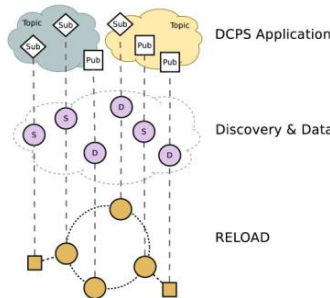


Fig. 9. RELOAD-based Discovery for DCPS.

DDS discovery– In DCPS, prior to sharing data, publishers and subscribers must discover each other (i.e., a rendezvous function is needed). For doing that, among others, a scalable solution is to distribute the discovery information among the members by using a P2P overlay. One of these solutions is Resource Location And Discovery (RELOAD), an IETF specification for building and maintaining P2P systems on the Internet. We propose a scalable RELOAD-based solution (Fig. 9) to the discovery problem in DCPS environments [24]. Our proposal demonstrates the extensibility of RELOAD by proposing a RELOAD usage that adapts it for performing rendezvous in DCPS-based deployments.

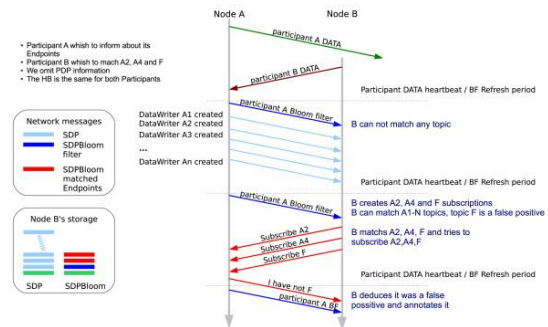


Fig. 10. SDPBloom for DDS Discovery.

To provide interoperability and transparency between different DDS implementations, any compliant DDS implementation must support at least the SDP (Simple Discovery Protocol). We have designed and evaluate SDPBloom, an SDP alternative –which uses Bloom Filters (BF)– (Fig. 10) that increases DDS scalability [25].

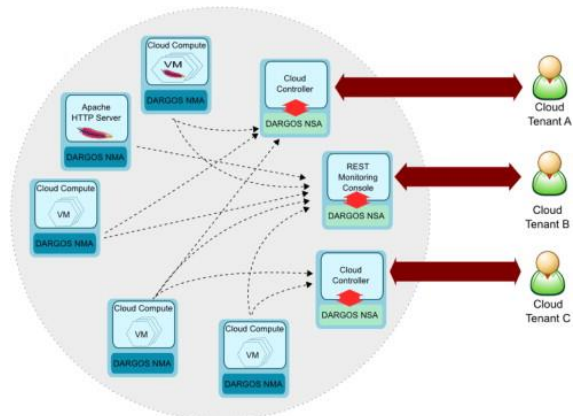


Fig. 11. DARGOS.

DDS for cloud monitoring– One of the most important features in Cloud environments is to know the status and the availability of the physical resources and services present in the current infrastructure. We have proposed [26][27] a Distributed Architecture for Resource management and monitoring in clouds (DARGOS) (Fig. 11), a completely distributed and highly efficient Cloud monitoring architecture to disseminate resource monitoring information based in DDS. DARGOS ensures an accurate measurement of physical and virtual resources in the Cloud keeping at the

same time a low overhead. In addition, DARGOS is flexible and adaptable and allows defining and monitoring new metrics easily.

Multimedia over DDS– DDS is a good candidate for multimedia (such as audio and video) content delivery. The rich set of QoS defined by the DDS standard have to be tuned to enable DDS to distribute multimedia data. Our research is aimed to determine the optimal QoS policies settings (Fig. 12) that can improve the multimedia experience in complex scenarios such as reduced bandwidth environments without the need of classical centralized services [28][29].



Fig. 12. DDS data-spaces allow sharing rich multimedia content.

DDS scalability – Initially DDS systems were restricted to a single, isolated, specific Domain, which was usually confined to a LAN and rarely exceeded one thousand computers. To overcome this restriction, we proposed the DDS Routing Service (DDS-RS) [30]. The DDS-RS (Fig. 13) is a generic service capable of transparently bridging DDS Domains as well as adapting among different data schemas. A key benefit of our approach is that the service can interoperate with existing DDS applications, bridging them without requiring them to be modified. We also showed that the impact of the service on the communications performance is well within the acceptable limits of most real-world uses of DDS. DDS-RS was adopted as a product by RTI DDS Connex © implementation.

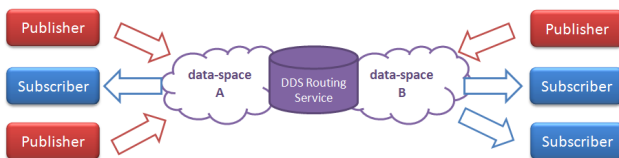


Fig. 13. The DDS-RS transparently bridges DDS domains.

III. LATEST RESEARCH PROJECTS AND CONTRACTS

Our research projects and contracts, both with public and private funding, may be found at [1]. In this section we summarize the latest projects and contracts that are related to our current research.

- Project TIN2013-46223 (subprogram TIN) of the Spanish Ministry of Science and Innovation, titled “Architecture

for 5G mobile communications based on Software Defined Networks,” 2014-2016.

- Project PYR-2014-18 “Virtualization of base stations for 5G mobile communications” of the Granada Excellence Network of Innovation Laboratories (GENIL [31]), 2014.
- Project PYR-2014-20 “Future multimedia networks based on SDN and quality of experience” of GENIL, 2014.
- Project TIN2010-20323 (subprogram TIN) of the Spanish Ministry of Science and Innovation, titled “Radio resource management for 3G Long Term Evolution radio access networks based on Quality of Experience criteria”, 2010-2013.
- Contract “Quality of Experience Enhancements in 3G Long Term Evolution Networks” between the University of Granada and Telefonica I+D, 2011.
- Contract “Framework Agreement Between INDRA Sistemas S.A. and the University of Granada for SESAR Program. Identification of Technology and Services Options & SWIN Prototype” between the University of Granada and INDRA SISTEMAS S.A., 2010-2012.
- Contracts related to DDS with Real-Time Innovations, Inc., 2007 – 2011.

IV. TEACHING

The members of WiMuNet impart the following subjects in the degrees of Telecommunications Engineering and Computer Science Engineering, as well as in their respective masters:

- Degree of Telecommunications Engineering:
 - o Wireless Networks and Mobility (http://dtstc.ugr.es/it/gitt_rim/)
 - o Multimedia Networks (http://dtstc.ugr.es/it/gitt_rm/)
 - o Switching Systems (http://dtstc.ugr.es/it/itt_sc/)
 - o Access and Corporate Networks (http://dtstc.ugr.es/it/gitt_rac/)
 - o Development of Network Applications (http://dtstc.ugr.es/it/gitt_dar/)
 - o Data Transmission and Computer Networks (http://dtstc.ugr.es/it/gitt_tdrc/)
 - o Communications Infrastructures and Networks (http://dtstc.ugr.es/it/itt_irc/)
 - o Telematics Laboratory (http://dtstc.ugr.es/it/gitt_lt/)
- Degree of Computer Science Engineering:
 - o Networking Fundamentals (http://dtstc.ugr.es/it/gii_fr/)
 - o Multiservice Networks (http://dtstc.ugr.es/it/ii_rm/)
- Master of Telecommunications Engineering:
 - o Advanced Networks and Services (http://dtstc.ugr.es/it/mit_rsa/)

o Internet Technologies and Distributed Systems

- Master of Computer Science Engineering:
 - o Computer Science Projects Planning and Management (http://dtstc.ugr.es/it/mii_pgpi/)

V. PH.D. THESIS

The following are the Ph.D. thesis carried out at the WiMuNet Lab or by WiMuNet members:

- Jonatan Prados-Garzon: "Architecture for 5G Mobile Networks based on SDN (*Software-Defined Networking*)," under development. Supervised by Prof. Lopez-Soler.
- Javier Povedano-Molina, "A data-centric publish/subscribe approach for multimedia service provisioning," 2013. Supervised by Prof. Lopez-Soler. Available at <http://hera.ugr.es/tesisugr/2249330x.pdf>
- Jose M. Lopez-Vega, "Discovery and signaling enhancements for data-centric publish-subscribe environments," 2013. Supervised by Prof. Lopez-Soler and Dr. Pardo-Castellote. Available at <http://hera.ugr.es/tesisugr/2220037x.pdf>
- Jorge Navarro-Ortiz, "Mechanisms for improving the Quality of Experience over IEEE 802.11 networks," 2010. Supervised by Prof. Lopez-Soler. Available at <http://hera.ugr.es/tesisugr/18858764.pdf>
- Erika P. Alvarez-Flores, "Selective packet dropping in active queue management mechanisms," 2009. Supervised by Prof. Lopez-Soler. Available at <http://hera.ugr.es/tesisugr/17983824.pdf>
- Juan J. Ramos-Munoz, "Improvements for voice transmission over IP considering experienced quality criteria," 2009. Supervised by Prof. Lopez-Soler. Available at <http://hera.ugr.es/tesisugr/17891875.pdf>
- Pablo Ameigeiras, "Packet scheduling and Quality of Service in HSDPA," 2003. Supervised by Prof. Preben E. Mogensen and Ph.D. Jeroen Wigard. Available at http://vbn.aau.dk/files/14366040/pablo-phdthesis_day28032004_time2300-2.pdf
- Juan M. Lopez-Soler, "Aproximación experimental al mejor compromiso R-D para la codificación de la voz a muy bajo bit-rate", 1995. Supervised by Prof. Antonio Rubio Ayuso.

VI. AWARDS

Awards granted to WiMuNet members are included next:

- Certificate of the Higher Technical School of Computer Science and Telecommunications Engineering to 4 members of the team recognizing the excellence of their research.
- IEEE article of the month (April 2014) [14].

- Best Paper 2013 of the Transactions on Emerging Telecommunications Technologies journal with the paper [13].
- Paper [13] selected as one of the top articles published on the Transactions on Emerging Telecommunications Technologies journal of 2013 (Editor's Pick 2013).
- M.Sc. thesis "Integration scenarios for satellite and terrestrial LTE networks for seamless video delivery," realized by Mr. Ramon Gigante Carrizosa and directed by Dr. Ameigeiras. First position in the 1st edition of M.Sc. thesis awards of the Dpt. of Signal Theory, Telematics and Communications of the University of Granada.
- M.Sc. thesis "Performance evaluation of MIMO schemes in closed-loop for wireless networks: signaling and link adaptation," realized by Mr. Juan Jose Lopez Villar and supervised by Dr. Ameigeiras and Jonathan Prados Garzon. Second position (shared) in the 1st edition of M.Sc. thesis awards of the Dpt. of Signal Theory, Telematics and Communications of the University of Granada.
- M.Sc. thesis "Performance evaluation of MIMO schemes in closed-loop for wireless networks: channel and receptor modelling," realized by Mrs. Alba Maria Rivera Perez and supervised by Dr. Ameigeiras and Jonathan Prados Garzon. Second position (shared) in the 1st edition of M.Sc. thesis awards of the Dpt. of Signal Theory, Telematics and Communications of the University of Granada.
- M.Sc. thesis "Femtocell deployment in multioperator environments over TV White Spaces," realized by Ms. Ana Maria Lopez Perez and supervised by Dr. Navarro-Ortiz. First position in the 2nd edition of M.Sc. thesis awards of the Dpt. of Signal Theory, Telematics and Communications of the University of Granada.
- M.Sc. thesis "Video tester," realized by Mr. Inaki Ucar Marques and supervised by Prof. Lopez-Soler and Dr. Navarro-Ortiz. Finalist of the ONO award given by the Spanish Official Association of Telecommunications Engineers. "Highly Commended Award" in the category of M.Sc. Thesis with Best Business Perspectives in ICT from the Higher Technical School of Computer Science and Telecommunications Engineering of the University of Granada.
- M.Sc. thesis "YouTube mobile traffic modeling," realized by Hector A. Aguilera-Garcia and directed by Dr. Ramos-Munoz. Secondary award in the category of M.Sc. Thesis with Best Business Perspectives in the ICT industries from the ETSIIT.
- Juan J. Ramos-Munoz was awarded with the "Premio extraordinario de fin de carrera," prize awarded to the most outstanding graduate in the Computer Science degree, University of Granada, 2001.

- Juan J. Ramos-Munoz was awarded with the “Premio extraordinario de fin de carrera,” in the Computer Systems degree, University of Granada, 1999.

VII. PATENTS

The members of the team hold the following patents:

- P. Ameigeiras, J. Navarro-Ortiz, J.M. Lopez-Soler, R. Garcia-Perez, Q. Perez-Tarrero, J. Lorca-Hernando, “A method and a system for scheduling the downlink in Long Term Evolution (LTE) networks based on Quality of Service (QoS),” applicant Telefonica S.A., spanish patent ES2429663 A2, international patent PCT/EP2013/053324, 2013.
- P. Ameigeiras, E. Sillasto, J. Wigard, “Adaptive delayed ACK switching for TCP applications,” applicant Nokia Corporation, US Patent 7,225,266 B2, 2007.
- P. Ameigeiras, J. Wigard, “Method, system and transmitting side protocol entity for sending packet data units for unacknowledged mode services,” applicant Nokia Corporation, international patent PCT/FI2004/000700, 2005.
- J. Wigard, T. Kolding, P. Mogensen, P. Ameigeiras, W. Hwang, H. Hakkinen, “Scheduling retransmission in Radio Access Networks,” applicant Nokia Corporation, international patent PCT/IB2003/005716, 2004.
- A. Tölli, J. Navarro-Ortiz, H. Montes-Linares, “A Traffic Management Method that Divides Time Slots into Sub-Blocks for Real Time and Non-Real Time Traffic,” applicant Nokia Corporation, international patent PCT/FI2004/000285, 2004.
- P. Ameigeiras, E. Sillasto, J. Wigard, “Method and system for dispatching multiple TCP packets from communication systems,” applicant Nokia Corporation, international patent PCT/FI2001/001054, 2003.

VIII. PEOPLE

This section includes a summary of the curriculum vitae of the members of WiMuNet.

A. Juan M. Lopez-Soler



Juan M. Lopez-Soler is Full Professor at the Department of Signals, Telematics and Communications of the University of Granada (Spain). He teaches undergraduate and graduate courses in Computer Networks, Data Transmission, and Multimedia Networking. He received the B.S. degree in Physics (Electronics). In 1991-92, he joined the Institute for Systems Research (formerly SRC) at the University of Maryland (USA) as Visiting Faculty Research Assistant. In 1995 he received the Ph.D. degree from the University of Granada. He has participated in more than 15 public and private funded research projects, being the coordinator in 7 of them. He has published more than 50 papers in international conferences and journals. His research

interests include wired and wireless multimedia networking and real-time middleware.

B. Pablo Ameigeiras



Pablo Ameigeiras received his M.Sc.E.E. degree in 1999 by the University of Malaga, Spain. He carried his Master Thesis at the Chair of Communication Networks, Aachen University (Germany). In 2000, he joined the Cellular System group at the Aalborg University (Denmark) where he carried out his Ph.D. thesis. After finishing his Ph.D., Pablo worked in Optimi (now Ericsson). In 2006, he joined the Department of Signal Theory, Telematics, and Communications at the University of Granada (Spain). Since then, he has been leading several projects in the field of LTE and LTE-Advanced systems. Currently, his research interests include 4G and 5G wireless systems.

C. Jorge Navarro-Ortiz



Jorge Navarro Ortiz is Associate Professor (*Profesor Contratado Doctor*) at the Department of Signal Theory, Telematics and Communications of the University of Granada. He obtained his M.Sc. in Telecommunications Engineering by the University of Malaga (Spain) in 2001. Afterwards, his professional career was developed in the mobile communications field, working at Nokia Networks, Optimi (acquired by Ericsson) and Siemens. He started working at the University of Granada in 2006, where he got a Ph.D. in Telecommunications Engineering with the thesis titled “Mechanisms for Improving the Quality of Experience over IEEE 802.11 Networks”, realized under the supervision of Prof. Lopez-Soler. During this work, he performed a research stay at the *Dipartimento di Ingegneria dell’Informazione* of the University of Pisa in summer 2009. During summer 2012 he was a visiting researcher with the Broadband Wireless Networking Lab of the School of Electrical and Computer Engineering of the Georgia Institute of Technology. His research interests include cognitive radio, heterogeneous networks, 4G and 5G wireless systems among others.

D. Juan J. Ramos-Munoz



Juan J. Ramos-Munoz received his M.Sc. in Computer Sciences in 2001 by UGR (University of Granada, Spain). Since 2009 he holds a doctorate degree from UGR. He is a lecturer in the Department of Signals Theory, Telematics and Communications of UGR. He is also member of the Wireless and Multimedia Networking Lab. His research interests are focused on software defined networks, and 5G, real-time multimedia streaming, quality of experience (QoE) assessment, Distributed Data Service (DDS), and adaptive mechanisms for providing QoS to Real-Time multimedia flows.

E. Jonathan Prados-Garzon



Jonathan Prados-Garzon received his MSc. in Telecommunications Engineering by the University of Granada (Spain) in 2011. He was granted a FPU fellowship by the Spanish Ministry of Education on 2014 and started his Ph.D. studies at the Department of Signal Theory, Telematics and Communication of the University of Granada. His research interests are focused on Quality of Experience (QoE) and 3G LTE systems.

IX. CONCLUSIONS

This paper summarizes our main research and teaching activities, and presents the members of the WiMuNet Lab of the University of Granada.

ACKNOWLEDGMENTS

This work is supported by the Spanish Ministry of Science and Innovation (project TIN2013-46223).

REFERENCES

[1] WiMuNet web page, available at <http://wimUNET.UGR.ES> (old version available at <http://dtstc.UGR.ES/tl/home.html>).

[2] Andalusian Scientific Information System (SICA), available at <https://sica2.cica.es/>.

[3] P. Ameigeiras, J.J. Ramos-Munoz, J. Prados-Garzon, J. Navarro-Ortiz, J.M. Lopez-Soler, "Link-level Access Cloud architecture design based on SDN for 5G Networks," IEEE Network, vol. 29, issue 2, pp. 24-31, March 2015.

[4] A.F. Cattoni, P.E. Mogensen, S. Vesterinen, M. Laitila, L. Schumacher, P. Ameigeiras, J.J. Ramos-Munoz, "Ethernet-based mobility architecture for 5G," IEEE 3rd International Conference on Cloud Networking (CloudNet 2014), pp. 449-454, 2014.

[5] P. Ameigeiras, D.M. Gutierrez-Estevez, J. Navarro-Ortiz, "Dynamic deployment of small cells in TV White Spaces," IEEE Transactions on Vehicular Technology, issue 99, October 2014.

[6] P. Ameigeiras, Y. Wang, J. Navarro-Ortiz, P.E. Mogensen, J.M. Lopez-Soler, "Traffic models impact on OFDMA scheduling design," Eurasip Journal on Wireless Communications and Networking, 2012.

[7] P. Ameigeiras, J.J. Ramos-Munoz, J. Navarro-Ortiz, J. Ramiro-Moreno, J.M. Lopez-Soler, "QoE evaluation of scheduling algorithms for NRT services in LTE," ICT-MobileSummit 2009, Santander (Spain), June 2009.

[8] P. Ameigeiras, J.J. Ramos-Munoz, J. Navarro-Ortiz, P. Mogensen, J.M. Lopez-Soler, "QoE oriented cross-layer design of a resource allocation algorithm in Beyond 3G systems," Computer Communications, March 2010.

[9] J.J. Ramos-Munoz, L. Yamamoto, C. Tschudin, "Serial experiments online," Computer Communication Review, vol. 38, 2, pp. 31-42, 2008.

[10] J.J. Ramos-Munoz, A.M. Gomez-Garcia, J.M. Lopez-Soler, "Intelligibility evaluation of a VoIP multi-flow block interleaver," Lecture Notes in Computer Science, pp. 200-205, 2005.

[11] J.J. Ramos-Munoz, J.M. Lopez-Soler, "Low delay multiframe block interleavers for real-time audio streaming," Lecture Notes in Computer Science, vol. 3420, pp. 909-916, 2005.

[12] E.P. Alvarez-Flores, J.J. Ramos-Munoz, P. Ameigeiras, J.M. Lopez-Soler, "Selective packet dropping for VoIP and TCP flows," Telecommunication Systems, 46 (1), pp. 1-16, 2011.

[13] P. Ameigeiras, J.J. Ramos-Munoz, J. Navarro-Ortiz, and J.M. Lopez-Soler, "Analysis and modeling of YouTube traffic," Transactions on Emerging Telecommunications Technologies, vol. 23, issue 4, pp. 360-377, June 2012.

[14] J.J. Ramos-Munoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, J.M. Lopez-Soler, "Characteristics of mobile YouTube," IEEE Wireless Communications Magazine, special issue on Mobile Social Networks, February 2014.

[15] P. Ameigeiras, A. Azcona-Rivas, J. Navarro-Ortiz, J.J. Ramos-Munoz, J.M. Lopez-Soler, "A simple model for predicting the number and duration of rebuffering events for YouTube flows," IEEE Communications Letters, vol. 16, issue 2, pp. 278-280, 2012.

[16] J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, J.M. Lopez-Soler, "Simulation-based performance study of YouTube service in 3G LTE," IEEE 14th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2013), Madrid (Spain), June 2013.

[17] OpenDayLight, available at <http://www.opendaylight.org/>.

[18] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, RFC 3550, "RTP: A Transport Protocol for Real-Time Applications," 2003.

[19] J. Navarro-Ortiz, P. Ameigeiras, J.M. Lopez-Soler, J. Lorca-Hernando, Q. Perez-Tarrero, R. Garcia-Perez, "A QoE-Aware Scheduler for HTTP Progressive Video in OFDMA Systems," IEEE Communications Letters, April 2013.

[20] J. Navarro-Ortiz, G. Stea, J.M. Lopez-Soler, "Quality of Experience Based Resource Sharing in IEEE 802.11e HCCA," European Wireless 2010, Lucca (Italia), April 2010.

[21] J. Navarro-Ortiz, P. Ameigeiras, J.J. Ramos-Munoz, J.M. Lopez-Soler, "An 802.11e HCCA scheduler with an end-to-end quality aware territory method," Computer Communications, vol. 32, issue 11, special issue on Heterogeneous Networking for Quality, pp. 1281-1297, 2009.

[22] J. Navarro-Ortiz, P. Ameigeiras, J.J. Ramos-Munoz, J.M. Lopez-Soler, "Removing redundant TCP functionalities in wired-cum-wireless networks with IEEE 802.11e HCCA support," International Journal of Communication Systems, 2013.

[23] Real-Time Innovations Inc., available at <http://www.rti.com>.

[24] J.M. Lopez-Vega, G. Camarillo, J. Povedano-Molina, J.M. Lopez-Soler, "RELOAD extension for data discovery and transfer in data-centric publish subscribe environments," Computer Standards and Interfaces, vol. 36, issue 1, pp. 110-121, 2013.

[25] J. Sanchez-Monedero, J. Povedano-Molina, J.M. Lopez-Vega, J.M. Lopez-Soler, "Bloom filter-based discovery protocol for DDS middleware," Journal of Parallel and Distributed Computing, vol. 71 (n° 10), pp. 1305-1317, Oct. 2011.

[26] J. Povedano-Molina, J.M. Lopez-Vega, J.M. Lopez-Soler, A. Corradi, L. Foschini, "DARGOS: A highly adaptable and scalable monitoring architecture for multi-tenant clouds," Future Generation Computer Systems, vol. 29, issue 8, pp. 2041-2056, Oct. 2013.

[27] A. Corradi, L. Foschini, J. Povedano-Molina, J.M. Lopez-Soler, "DDS-enabled cloud management support for fast task offloading," 17th IEEE Symposium on Computers and Communication, Capadocia (Turkey), Jul. 2012.

[28] J. Povedano-Molina, J.M. Lopez-Vega, J.M. Lopez-Soler, "EMDS: an Extensible Multimedia Distribution Service," Workshop on Real-time, Embedded and Enterprise-Scale Time-Critical Systems, Arlington, VA (USA), May 2010.

[29] J.M. Lopez-Vega, J. Sanchez-Monedero, J. Povedano-Molina, J.M. Lopez-Soler, "QoS policies for audio/video distribution over DDS," Workshop on Distributed Object Computing for Real-Time and Embedded Systems, Arlington, VA (USA), Jul. 2008.

[30] J.M. Lopez-Vega, J. Povedano-Molina, G. Pardo-Castellote, J.M. Lopez-Soler, "A content-aware bridging service for publish/subscribe environments," Journal of Systems and Software, vol. 86, issue 1, pp. 108-124, Jan. 2013.

[31] Granada Excellence Network of Innovation Laboratories (GENIL) project, available at <http://genil.UGR.ES>.

Maqueta para la evaluación de la calidad de experiencia del servicio videostreaming 3D en LTE

Inmaculada Palomo, Marta Solera Delgado, Matías Toril, Javier Poncela-González, Gerardo Gómez.

{inmapalomo,msolera,mtoril,javier.gomez}@ic.uma.es

Dpto. de Ingeniería de Comunicaciones. ETSI Telecomunicación. Universidad de Málaga.

Campus de Teatinos s/n 29071 Málaga (España)

Abstract-With the deployment of the first commercial LTE (*Long Term Evolution*) networks, mobile operators need to understand how QoS (*Quality of Service*) network and codec parameters affect the subjective quality of high-definition and 3D video, as perceived by customers. In this article, we describe the development of a testbed system for evaluating the QoE (*Quality of Experience*) of 3D Video Streaming Service over LTE. The proposed system consists of three elements: the streaming server, the IP mobile network emulator and the streaming client. In the testbed, different network conditions are configured by setting network emulator parameters based on the results obtained by a system-level LTE simulator. Results show how average network load and user position inside the cell can have a strong impact on the QoS and QoE perceived by the end 3D video user.

I. INTRODUCCIÓN

En los últimos años, la industria de la electrónica de consumo ha dedicado numerosos esfuerzos a llevar el vídeo y la televisión tridimensional (3D) a los hogares. En paralelo, las industrias del cine y del deporte también han apostado por las tres dimensiones con la creación de contenidos 3D atractivos para el espectador. Así, un estudio de CISCO, principal fabricante de dispositivos de red, prevé que en 2017 el tráfico de vídeo de alta definición y 3D alcance el 63% del tráfico total en Internet.

Lamentablemente, las redes móviles tradicionales no están diseñadas para la distribución de vídeo de alta definición y 3D. Para que estos formatos se consoliden, es preciso distribuir de forma eficiente la gran cantidad de información que requieren ambos formatos. Por ello, se hacen necesarias nuevas formas de codificar, comprimir y transportar este tipo de datos.

Para evaluar el rendimiento de sus redes, los operadores de telefonía móvil consideraban hasta hace poco medidas objetivas de calidad de servicio (*Quality of Service*, QoS), como el retardo o la velocidad de transmisión. No obstante, aunque estas medidas están relacionadas con el grado de satisfacción del usuario, la relación exacta no es directa. Por ello, los operadores están cambiando la estrategia clásica de gestionar la red, centrada en el control del rendimiento de la red, por una estrategia que ponga el foco en satisfacer al usuario.

En este contexto, el reciente despliegue de las redes móviles comerciales de cuarta generación (4G), basadas en el estándar LTE (*Long Term Evolution*), requiere la realización de estudios que determinen el impacto real de los parámetros de QoS y de la compresión en la experiencia del servicio de distribución de vídeo 3D en este tipo de redes.

En trabajos anteriores, se ha estudiado la calidad de experiencia en la distribución de vídeo 3D sobre redes IP cableadas [1] [2] e inalámbricas [3]. Otros trabajos han estudiado la distribución de vídeo convencional 2D en redes móviles [4]. Los autores, en un artículo previo [5], presentan una plataforma de sistema de distribución de vídeo 3D en tiempo real sobre IP. Con dicha maqueta, se puede evaluar con usuarios reales el impacto de los parámetros de calidad de servicio definidos en la red (p.ej., retardo, tasa de pérdidas) y los parámetros configurables a nivel de aplicación (p.ej., codificador, tipo de contenido) sobre la calidad de experiencia (*Quality of Experience*, QoE) percibida por el observador. Su principal novedad frente a maquetas similares descritas en la bibliografía (p.ej., [6] [7]) es la combinación del tráfico de vídeo 3D con la emulación de una red celular. En ese artículo se presentaba un ejemplo sencillo de configuración de la maqueta para evaluar la calidad del servicio de *videostreaming* 3D en una red LTE.

Este trabajo complementa al anterior, profundizando en la configuración de los elementos que conforman la maqueta. Por un lado, se presentan nuevas modificaciones en el emulador de red para modelar mejor las características del retardo de paquetes cuando el proceso de llegada de paquetes es discontinuo en el tiempo, como es el caso del servicio de *videostreaming*. Por otro lado, se describe una nueva metodología de ajuste de los parámetros del emulador de red basándose en los resultados obtenidos con un simulador de nivel de sistema. Una vez configurado el emulador, se evalúan 4 escenarios diferentes que pretenden reflejar la experiencia de un usuario del servicio de *videostreaming* 3D cuando se sitúa en el centro o en el borde de la celda, en condiciones de baja o alta carga de red.

El resto del artículo se estructura de la siguiente manera. En la sección II se hace una breve introducción a la maqueta de emulación de red LTE para el servicio de *videostreaming*, describiendo la configuración de las herramientas utilizadas. La sección III explica la simulación de la interfaz radio LTE. La sección IV cuenta la metodología experimental de prueba del sistema. La sección V muestra los resultados y, por último, la sección VI expone las principales conclusiones.

II. SISTEMA DE EMULACIÓN DE VIDEOSTREAMING EN LTE

Existen diferentes metodologías para evaluar el comportamiento extremo a extremo de una red de telecomunicación. La primera de ellas es la simulación, que utiliza un modelo simplificado de un escenario real y que proporciona unos resultados limitados a la complejidad del modelo. Una segunda alternativa es la emulación, que, dados

unos indicadores de rendimiento generales de la red, imita su comportamiento en un entorno controlado. La tercera alternativa es la realización de medidas de campo, que se toman directamente de los elementos de la red en un escenario real. La plataforma descrita en este trabajo combina todas estas estrategias, empleando un emulador al que se le introducen como parámetros de entrada los valores de los indicadores de red (e. g., retardo, *jitter*, porcentaje de pérdidas de paquetes) extraídos de simulaciones de nivel de red y/o medidas de campo de una red LTE.

La maqueta propuesta consta de tres elementos: el servidor de *streaming*, el emulador de red IP móvil y el cliente de *streaming*. En los siguientes apartados se describen los formatos de vídeo 3D, los protocolos de red y los elementos que forman la maqueta.

A. Formatos de vídeo 3D

El formato de representación de vídeo 3D más simple es el formato CSV (*Conventional Stereo Video*). El formato CSV se compone de dos flujos de vídeo independientes (pero sincronizados), correspondientes al ojo izquierdo y derecho de un hipotético observador. Ambos flujos se pueden distribuir en paralelo utilizando sistemas convencionales, estrategia conocida como *Simulcast*. Este formato no requiere ningún procesamiento de geometría espacial, aportando sensación de profundidad al vídeo 2D, pero a costa de doblar la tasa de transferencia de un flujo de vídeo convencional. Por ello, se hace imperativa la utilización de métodos de compresión de la información para adaptarla a las limitadas capacidades de la red y los terminales móviles.

En la plataforma desarrollada, se utilizan los formatos 3D de tasa de transferencia reducida más habituales, que son SbS (*Side-by-Side*) y MVC (*Multi-View Coding*). El formato SbS es un formato compatible con los sistemas de distribución de vídeo tradicionales. Para ello, se aplica un proceso de diezmado en el dominio espacial de ambos ojos y las imágenes diezmadas se unen en un único fotograma convencional. El formato MVC, utilizado en el soporte de almacenamiento Blu-ray 3D, aplica codificación diferencial entre ambas vistas para eliminar la información redundante entre ellas [3].

B. Distribución del servicio de streaming de vídeo

Para la distribución de vídeo en redes IP, se pueden usar diferentes combinaciones de protocolos de aplicación, como HTTP (*Hyper-Text Transfer Protocol*) o RTP (*Real Time Protocol*), y protocolos de transporte, como TCP (*Transport Control Protocol*) o UDP (*User Datagram Protocol*). La combinación HTTP-TCP, con ciertas modificaciones sobre los estándares, es la escogida para el *streaming* de descarga progresiva, utilizado, p.ej., por YouTube. Por el contrario, RTP-UDP se usa para videoconferencia (p.ej., *Skype*), por ser más eficiente en el uso del ancho de banda.

La Fig. 1 muestra la arquitectura de protocolos del sistema de evaluación propuesto. En la figura, se identifican el cliente, el emulador de red y el servidor. Tanto el cliente como el servidor soportan las diferentes combinaciones de protocolos, aunque en este trabajo se considera la opción del servicio de *videostreaming* 3D sobre RTP/UDP. Por su parte,

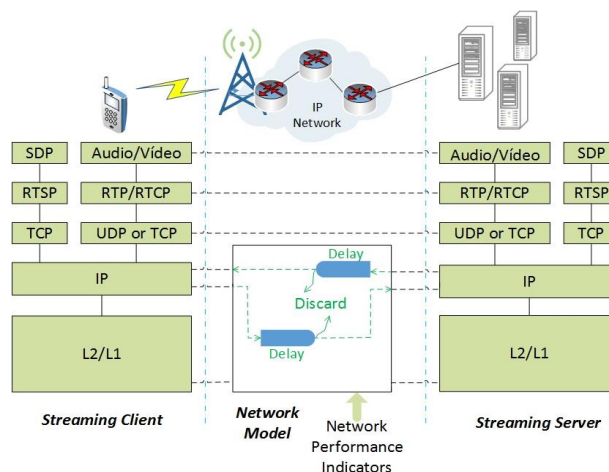


Fig. 1. Arquitectura de protocolos del sistema de evaluación.

el emulador de red introduce pérdidas y retardo sobre los paquetes, según su configuración de parámetros.

El servicio de *streaming* consiste en enviar los datos de forma continua estableciendo una conexión permanente entre cliente y servidor mediante RTSP (*Real-Time Streaming Protocol*). Una vez iniciada la sesión, normalmente los datos del vídeo y el audio serán enviados en sendas parejas de puertos RTP/RTCP (datos y control) sobre UDP, aunque también existe el mecanismo llamado *interleaving* que hace que RTSP se encargue del transporte de los datos.

Los indicadores de rendimiento principales para evaluar la calidad de la comunicación en servicios de vídeo son: velocidad de transferencia, retardo y *jitter*. La diferencia de retardo entre dos paquetes i y j , $D(i,j)$, y el *jitter* entre llegadas, $J(i)$, sufridos por cada uno de los flujos RTP se calculan a partir de las ecuaciones definidas en la RFC 3550,

$$D(i,j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \quad (1)$$

Donde S_i y S_j son los RTP *timestamp* de los paquetes i y j , y R_i y R_j son los instantes de llegada de los paquetes i y j en unidades de RTP *timestamp*, y

$$J(i) = J(i-1) + (|D(i-1,i)| - J(i-1))/16 \quad (2)$$

donde $J(i-1)$ es el *jitter* del paquete $i-1$ y $D(i-1,i)$ es la diferencia de tiempos de transmisión y recepción de los paquetes i e $i-1$ en orden de llegada (no necesariamente consecutivos).

C. Elementos de la maqueta

La Fig. 2 muestra los elementos hardware que conforman la maqueta del sistema, compuesta por un servidor, un emulador de red y dos tipos de cliente. Como elementos software se han utilizado VLC, *Wireshark* y *NetEm*.

El servidor es un PC con procesador i5-750 a 2.67GHz y 4 GB de RAM con sistema operativo Ubuntu 14.04 LTS 64 bits y *kernel* 3.11.10.4. en el que se ha instalado el software VLC *media player* como servidor de *streaming*.

Un primer tipo de cliente es un PC con las mismas características que el anterior en el que también se ha instalado el sistema operativo Ubuntu 14.04 LTS 64 bits. En



Fig. 2. Componentes de la maqueta de evaluación.

él, se utiliza también el mismo reproductor multimedia VLC, pero ahora como cliente de *streaming*. Tanto en el servidor como en este cliente se ejecuta el software *Wireshark* como analizador de tráfico.

El otro tipo cliente es un terminal móvil modelo LG Optimus 3D-P920 con sistema operativo *Android 2.2.2*, en el que se ha instalado *VLC media player para Android v0.1.3*. Éste se emplea para las medidas de calidad subjetivas sobre pantalla pequeña. Estas medidas pueden realizarse en tiempo real haciendo uso de la conectividad Wifi del terminal, o en tiempo diferido reproduciendo los archivos almacenados en el cliente PC con las secuencias de vídeo deterioradas tras su paso por el emulador de red.

El emulador de red se instala en un PC con un procesador i5-3470 a 3.2GHz, 8 GB de RAM con sistema operativo Ubuntu 12.04 LTS 64 bits. Este ordenador consta de dos tarjetas de red para conectar al servidor y al cliente de *streaming* a través de él. La conexión entre el servidor de *streaming* y el emulador y entre este y el cliente se hace por medio de un cable de par trenzado cruzado para tener una red completamente aislada. Se han de configurar convenientemente las tablas de enrutamiento a través del comando de Linux *ip route*. El software que realiza las funciones de emulador es *NetEm*.

La versión de *VLC media player* elegida ha sido v2.1.4 *Rincewind*. Este reproductor está muy extendido por su gratuidad, facilidad de uso, variedad de *codecs* soportados, rapidez, fiabilidad y versatilidad. Es una plataforma de código libre, que soporta la mayoría de formatos multimedia, así como varios protocolos de *streaming*. En nuestro caso, además de para la reproducción de archivos multimedia, se ha hecho uso de la captura y almacenamiento de la secuencia de reproducción en el terminal cliente, para su posterior procesamiento y análisis comparativo del vídeo emitido con el recibido.

También *VLC* permite incluir fácilmente nuevos *codecs* de vídeo. Así que se ha ampliado esta herramienta y se han incluido las rutinas JMVC, que es el software de referencia del códec MVC desarrollado por el JVT (*Joint Video Team*) [2].

Como aplicación para la captura y análisis de tráfico se ha escogido *Wireshark v1.6.7-1* por ser el analizador de protocolos más utilizado por la comunidad científica para realizar análisis y diagnosticar problemas en redes de comunicaciones. Proporciona una funcionalidad similar a la de *tcpdump*, sin embargo añade una interfaz gráfica y muchas opciones de organización y filtrado de capturas que hacen que sea un entorno más amigable.

Como emulador de red se ha escogido *NetEm*, el paquete *iproute2 v3.10.0-Iubuntu1*. Este software es un emulador de red de código abierto ampliamente utilizado en sistemas de evaluación y desarrollo de protocolos de red y aplicaciones. Reproduce la dinámica de una red mediante la introducción de retardos, jitter y pérdidas, duplicando o eliminando paquetes.

A partir de estos elementos hardware y software se construye la maqueta que nos permitirá emular un sistema de *videostreaming* sobre una red móvil. Es necesario obtener los parámetros que nos permitan modelar una red LTE para introducirlos de forma adecuada en el emulador de red.

D. El emulador de Red: NetEm

NetEm[8] es un emulador de red que permite controlar el retardo, las pérdidas y la duplicación y el reordenamiento de los paquetes IP. A partir de la versión 2.6.8 (2.4.28), esta herramienta se encuentra integrada en el kernel de Linux y se controla a partir del comando 'tc' que es parte del paquete *iproute2*. *NetEm* se configura a través de este comando y puede automatizarse mediante el uso de scripts en shell para mayor comodidad.

En *NetEm*, la pérdida de paquetes se implementa descartando al azar un porcentaje de los paquetes antes de que se pongan en cola. Dicha pérdida se especifica en la interfaz de comando como un porcentaje. La emulación de pérdidas de paquetes a ráfagas se realiza mediante uno de los modelos de pérdidas disponibles (correlación, modelo de cuatro estados, modelo Gilbert-Eliot).

Los parámetros de retardo y su variación, *jitter*, se describen por el valor medio (μ), la desviación estándar (σ) y, opcionalmente, un valor de correlación (ρ). Por defecto, *NetEm* utiliza una distribución uniforme de retardo y jitter, que puede cambiarse a otra función de distribución (pareto, paretonormal o normal) u otras específicas a partir de datos experimentales. Esta última opción es la elegida para configurar convenientemente el emulador a partir de los datos obtenidos del simulador LTE.

Aunque *NetEm* es una herramienta con mucho potencial, tiene defectos en su comportamiento, como evidencian trabajos anteriores [9]. Estos estudios muestran que si a *NetEm* se le obliga a guardar la integridad del flujo de paquetes, para evitar su reordenamiento, se modifica el valor medio y la forma de la función de distribución del retardo. Este comportamiento hace que *NetEm*, en su configuración habitual, no sea apropiado para modelar sistemas de colas.

En este trabajo, para evitar este comportamiento, se modifica el código fuente de *NetEm*. En concreto, se modifica el método que se encarga de encolar los paquetes (función *netem_enqueue*). El objetivo del cambio es modelar la acumulación de retardo que se produce cuando los paquetes llegan juntos al principio de cada cuadro.

Para ello, el retardo de transmisión de paquete se calcula como la suma de dos componentes, llamados retardo en cola y retardo de transmisión como primero en cola. El retardo en cola, D_q , es la diferencia de tiempo entre el instante de llegada del paquete y el instante en el que paquete pasa a ser el primero en la cola. El retardo de transmisión como primero en cola, D_{1st} , es la diferencia de tiempo entre el instante en el que el paquete pasa a ser el primero en la cola y el instante en que finalmente es servido.

```
static int netem_enqueue(structsk_buff *skb,
structQdisc *sch)
{
delay = tabledist(q->latency, q->jitter,
&q->delay_cor, q->delay_dist);

[...]

if (last) {
/*Last packet in queue is reference (now),
*calculate this time bonus and add
*from delay*/

delay += netem_skb_cb(last)->time_to_send - now;
delay = max_t(psched_tdiff_t, 0, delay);
now = netem_skb_cb(last)->time_to_send;
}
[...]
}
```

Fig. 3. Líneas de código añadidas en *NetEm*.

El retardo en cola de un paquete se calcula acumulando el retardo de transmisión de todos los paquetes que aún quedan por transmitirse en la cola. Por su parte, el retardo de transmisión primero en cola se genera de forma aleatoria a partir de la distribución de probabilidad obtenida en el simulador de red. Esta modificación en el código de *NetEm* evita el desordenamiento de los paquetes, haciendo que se transmitan según una cola con disciplina de servicio FIFO. Así se consigue la integridad del *videostreaming*, manteniendo los valores estadísticos y la forma de la función de distribución del retardo.

La Fig. 3 muestra las líneas de código que se introducen. En primer lugar, se obtiene el valor de retardo primero en cola a partir de la tabla que modela la distribución acumulada de probabilidad empírica (*Empirical Cumulative Density Function*, ECDF) del retardo (función *tabledist*). Después, si el paquete no encuentra la cola vacía (*last=1*), al retardo primero en cola se le añade el tiempo de espera en la cola (*netem_skb_cb(last)->time_to_send - now*). De esta manera, se consigue que no haya paquetes que se adelanten a otros debido a que su *delay* sea menor.

Para introducir la función de probabilidad de retardo obtenida desde el simulador LTE en el emulador, en la función *tabledist* ha sido necesario modificar el tipo de datos que soportaba inicialmente *NetEm*.

III. SIMULACIÓN DE INTERFAZ RADIO LTE

Para configurar los parámetros del emulador de red con valores realistas, se realiza una campaña de simulaciones con un simulador LTE de nivel de sistema [10]. El objetivo de estas simulaciones es estimar el rendimiento de la transmisión de paquetes en la red de acceso con distintas condiciones radio.

La Fig. 4 muestra el escenario, con 53 celdas con antenas omnidireccionales situadas en el centro de cada celda, distribuidas de forma regular en el espacio. En dicho escenario, se distribuyen varios usuarios por celda del servicio de *videostreaming* 3D. La tabla 1 describe los parámetros del simulador. Por simplicidad, se simula solo el enlace descendente, que es el principal cuello de botella en la descarga de secuencias de vídeo de un servidor.

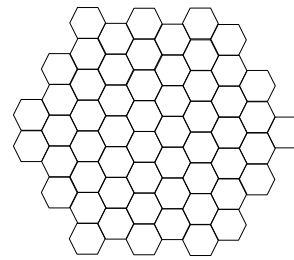


Fig. 4. Escenario de simulación.

Tabla 1. Parámetros del simulador de red LTE.

Frecuencia de portadora	2.5 GHz
Ancho de banda del sistema	10 MHz (50 Physical Resource Blocks)
Número de celdas	57
Radio de celda	1.5 km
Modelo de propagación	Hata
	Desvanecimiento lento lognormal 6dB
Potencia isotropa radiada efectiva	43 dBm
Densidad de potencia de ruido	-174 dBm/Hz
Scheduler eNB	Proportional Fair
Esquema de transmisión MIMO	2x2 Beamforming
Modelo de canal	Extended pedestrian A
Método decodificación de canal	Soft-Output Viterbi Algorithm
BLER objetivo	10%
Modelo de servicio	Traza de vídeo real H.264
Tiempo de red simulado	40s
Resolución temporal	1ms (TTI)

La salida principal obtenida del simulador es una traza con los retardos de transmisión de los paquetes de nivel IP en el enlace descendente obtenidas por un usuario en unas condiciones de red determinadas. Se distinguen 4 condiciones diferentes de trabajo, que pretenden reflejar la experiencia de un usuario cuando se sitúa en el centro o en el borde de la celda, en condiciones de baja o alta carga de red. La posición del usuario en la celda determina la calidad de señal (relación señal a interferencia más ruido) experimentada por el usuario, mientras que las condiciones de carga de red determinan el número de recursos radio disponible para el usuario. En los experimentos, la carga de red se controla modificando el número de usuarios por celda entre 2 usuarios (para el que se obtiene un porcentaje de ocupación de bloques de recursos físicos de un 25%) y 5 usuarios (para el que se obtiene un porcentaje del 75%). La posición de centro o borde celda se consigue fijando la posición en el escenario del usuario del que se obtiene la traza de retardo.

La Fig. 5 presenta la ECDF del retardo de transmisión de paquete para los 4 casos considerados, mientras que la tabla 2 detalla algunos estadísticos de la transmisión de paquetes. e aprecia cómo la mediana del retardo medio de transmisión de paquete oscila desde 8 ms para el caso más favorable (usuario centro de celda y carga baja) hasta valores por encima de los 200 ms para el caso más desfavorable (usuario borde celda y carga alta), que no aparece en la figura.

Para cada caso de uso, la traza de retardo de transmisión de paquete obtenida se procesa para obtener la traza correspondiente de retardo de paquete cuando el paquete es el primero en la cola. Concretamente, el retardo de paquete primero en cola, D_{1st} , se estima a partir del retardo total de transmisión, D , como

$$D_{1st}(i) = D(i) - D_q(i) = D(i) - \max(T(i-1) - A(i), 0) \quad (3)$$

donde D_q es el retardo de espera en cola, estimado a partir de la diferencia entre el instante de transmisión del paquete anterior, $T(i-1)$, y el instante de llegada del paquete en cuestión, $A(i)$.

Los valores de las componentes de retardo mostradas en la tabla 2 ponen de manifiesto que la mayor parte del retardo se debe al retardo de encolado. Un análisis detallado muestra que este comportamiento se debe a la llegada simultánea de paquetes al principio de cada cuadro. La Fig. 6 muestra un ejemplo de la evolución del tiempo entre llegadas de paquetes para una traza de vídeo. En ella, se aprecia que, tras un periodo inicial de intercambio de paquetes de señalización para establecer la sesión RTP, los paquetes que encapsulan el vídeo se generan en ráfagas que se repiten con periodo de cuadro (40 ms). En la figura se aprecia también que el número de paquetes por ráfaga mantiene un cierto patrón repetitivo, del que se puede deducir la estructura de grupo de imágenes del compresor, formada por 1 cuadro tipo I (*Intracuadro*) por cada 7 cuadros de tipo P (*Predictivo*). De la distancia entre picos de la serie de diferencias de instantes de llegada, se puede deducir que el número medio de paquetes que forman los cuadros I y P son 35 y 8, respectivamente.

A partir de las trazas de retardo obtenidas del simulador LTE se obtiene el retardo en cola que puede observarse en la Fig. 7. Este retardo presenta una curva en diente de sierra semiperiódico que muestra como al inicio de cada cuadro se produce un incremento drástico de este tiempo que va decreyéndose a medida que se sirven los paquetes.

También a partir de las trazas de retardo obtenidas del simulador LTE, se construye la ECDF del retardo de paquete primero en cola, D_{1st} . En la Fig. 8 se muestra la ECDF del retardo de paquete primero en cola para cada uno de los casos de usos. Se aprecia cómo la mediana de D_{1st} oscila entre 1 ms para el caso más favorable (usuario centro de celda y carga baja) hasta 6 ms para el caso más desfavorable (usuario borde celda y carga alta).

Con cada una de las ECDFs del retardo de paquete primero en cola de la Fig. 8, se genera una realización de 4096 valores aleatorios de retardo. Dichos valores, una vez convertidos a unidades de *tick* (1 *tick* = 64 μ s) y ordenados de menor a mayor, se introducen en la tabla de 512*8 valores utilizada por *Netem* para generar los retardos de paquete primero en cola en cada caso de uso.

Tabla 2. Estadísticos de transmisión de paquetes.

Carga de red (porcentaje)	Baja (25%)		Alta (75%)	
	Centro	Borde	Centro	Borde
Retardo total medio [ms]	14.98	48.82	30.92	4542.9
Retardo en cola medio [ms]	13.78	45.71	28.44	4258.8
Retardo primero en cola medio [ms]	1.20	3.11	2.48	284.1
Jitter medio [ms]	2.09	4.42	3.86	9.86

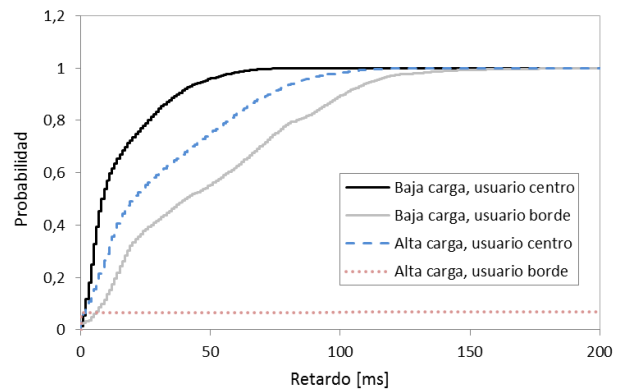


Fig. 5. Distribución acumulada de probabilidad del retardo de paquete total en distintas condiciones.

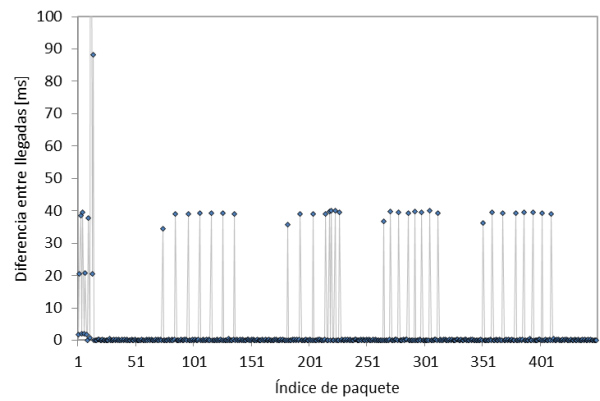


Fig. 6. Ejemplo de evolución del tiempo entre llegadas de paquetes.

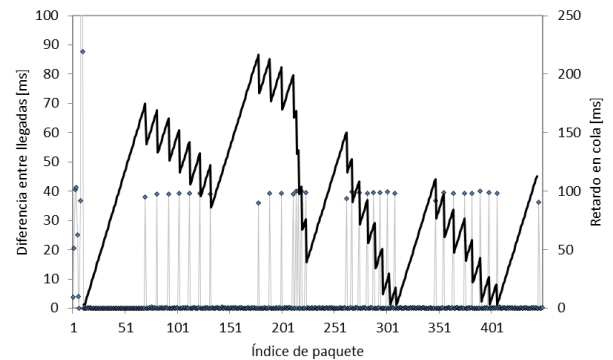


Fig. 7. Ejemplo de evolución del tiempo entre llegadas de paquetes y el tiempo de espera en cola que experimenta el usuario.

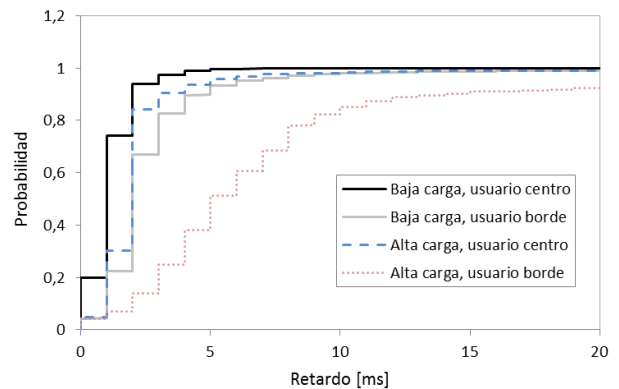


Fig. 8. Distribución acumulada de probabilidad del retardo de paquete primero en cola en distintas condiciones.

IV. METODOLOGÍA EXPERIMENTAL

Para validar el funcionamiento de la plataforma, primero se configuran los elementos de la maqueta según los resultados del simulador LTE, luego se lleva a cabo la distribución de secuencias 3D, y por último se procesan las medidas de retardo obtenidas.

La configuración de la maqueta consiste fundamentalmente en la selección de la secuencia de vídeo a distribuir, la selección de la tabla que modela la ECDF de retardo primero en cola en el emulador de red, y la configuración del tamaño del buffer del reproductor.

Antes de comenzar con la conexión entre cliente y servidor, se debe comprobar la sincronización de los equipos mediante el protocolo NTP (*Network Time Protocol*). Esta fase es muy importante, ya que de ello depende la bondad de los resultados. En la Fig. 9 puede observarse cómo la falta de sincronismo entre los equipos produce una deriva considerable, en este caso del orden de 1.43 ms por cada 1000 paquetes. En la Fig. 10 se observa el resultado de la misma simulación una vez que servidor y cliente han sido previamente sincronizados.

Como la frecuencia de trabajo de *NetEm* es directamente la marcada por el *kernel*, se ha modificado el tiempo de reloj del *kernel* a 1 ms para mejorar la resolución temporal de *NetEm*. Con el mismo fin, y para mejorar el sincronismo entre los tres PCs, se han habilitado los temporizadores de alta resolución del sistema operativo.

Una vez sincronizados los equipos, se lanzan los analizadores de tráfico en ambos equipos, emisor y receptor. Una vez activadas las herramientas de monitorización, se inicia la transmisión desde el servidor *VLC streaming Server* al cliente *VLC media player* de la secuencia de vídeo seleccionada en el cliente. Al concluir la reproducción, los datos capturados por los analizadores de tráfico del emisor y receptor se almacenan para su posterior análisis.

En la tabla 3 se detallan las características de las secuencias de vídeo empleadas y los valores de los parámetros de configuración de *NetEm* y *VLC*. Todas las secuencias de vídeo empleadas en las pruebas están codificadas en formato SbS. El tamaño del buffer se ha elegido de 50 ms, que es apropiado para un servicio interactivo de voz y vídeo cuyo retardo extremo a extremo esperado no debería ser mayor de unos 200 ms.

Para emular la red LTE, se configura el emulador de red *NetEm* de acuerdo a los cuatro escenarios que se han descrito en la sección anterior (alta/baja carga, usuario centro/borde). Para ello, se indica a *NetEm* la tabla concreta que modela el retardo primero en cola, generada a partir de una de las distribuciones de la Fig. 8.

Las trazas de retardo del cliente y servidor se capturan mediante *Wireshark*. Con su funcionalidad *RTP Analysis*, en teoría es posible medir el retardo medio, el *jitter* y las pérdidas de paquetes de un flujo de datos RTP. Sin embargo, durante los experimentos se ha comprobado que *Wireshark* no muestra los resultados de *jitter* y retardo medio de forma adecuada como cabría esperar, ya que no es capaz de interpretar correctamente la decodificación de flujos RTP. El retardo detectado es leído dentro de la *ExpertInfo* de *Wireshark*. Las estadísticas del *jitter* sufrido por el tráfico RTP se han calculado mediante la fórmula (2), a partir de los tiempos de transmisión, *timestamp* de la cabecera RTP del

paquete, y los tiempos de recepción de los paquetes capturados por el analizador del equipo receptor.

La calidad de experiencia ofrecida al observador final puede comprobarse reproduciendo las secuencias de vídeo recibidas en la pantalla del reproductor *VLC* del cliente. Para tener una medida objetiva del impacto de las fluctuaciones del retardo sobre la calidad del vídeo, se utiliza la funcionalidad de *VLC* que permite grabar las secuencias recibidas. Estas secuencias, una vez almacenadas, se procesan con *ffmpeg* [11], que permite evaluar la calidad de la secuencia en función del número de cuadros perdidos y duplicados.

Tabla 3. Parámetros de configuración de las pruebas.

Nombre de secuencia	pd01	pd02	pd03	pd04
Codec de vídeo	H.264/AVC 3086 kbps	H.264/AVC 2207 kbps	H.264/AVC 4385 kbps	H.264/AVC 1729 kbps
Duración [s]	48	23	40	60
Parámetros NetEm	[ECDF para cada uno de los 4 casos de uso]			
Parámetros VLC	Tamaño de buffer: 50ms			

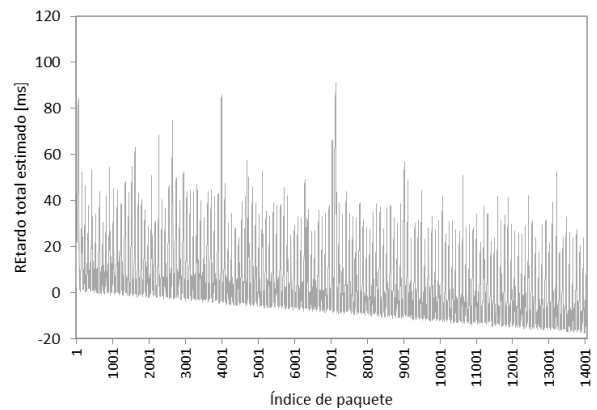


Fig. 9. Retardo extremo a extremo para cada paquete cuando el cliente y servidor no están sincronizados.

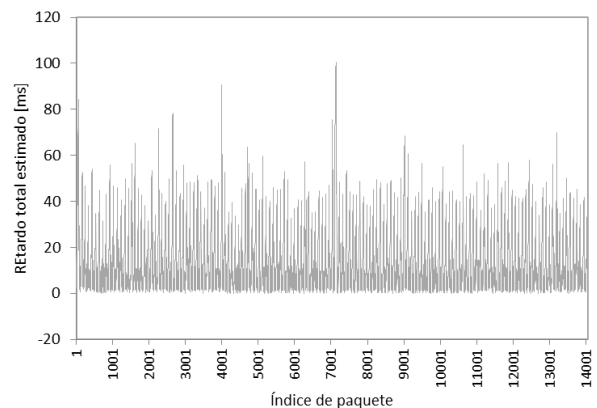


Fig. 10. Retardo extremo a extremo para cada paquete cuando el cliente y servidor están sincronizados.

V. RESULTADOS

En esta sección se muestran los resultados a los experimentos realizados para los cuatro casos de estudio que modelan las distintas condiciones que puede experimentar un

usuario, situado en el centro o en el borde de la celda, en condiciones de baja o alta carga de red. Por brevedad, solo se muestran los resultados obtenidos para una secuencia de vídeo (secuencia *pd01*, cuyas características se describen en la tabla 3). En la Fig. 11 se muestra un cuadro de la secuencia de vídeo escogida, donde se aprecian los fotogramas de la vista izquierda y derecha. En ella, se observa cómo esta secuencia consta de una única toma con poco movimiento, en la que se ubica una persona en primer plano gesticulando delante de un fondo estático.



Fig. 11. Imagen perteneciente a la secuencia *Bullinger* (secuencia *pd01*).

En la Fig. 12 se muestra la función de distribución acumulada del retardo extremo a extremo obtenida mediante el emulador de red (*NetEm*), comparada con la del simulador LTE, para cada uno de los cuatro casos. En la figura, la configuración del eje *x* ha sido seleccionada para apreciar con detalle las diferencias entre las curvas de menor retardo. Nótese que en el caso (carga alta, usuario de borde), al ser un sistema de cola inestable (es decir, tasa de llegada paquetes mayor que tasa de salida de paquetes), el retardo es muy elevado. Por ello, la curva correspondiente está desplazada en el eje *x*, no siendo visible. De la figura, puede concluirse que las curvas del simulador y el emulador se ajustan casi perfectamente para los casos (baja carga, usuario de centro) y (alta carga, usuario de centro), mientras que el ajuste es menos preciso, aunque aceptable, en el caso de (baja carga, usuario de borde).

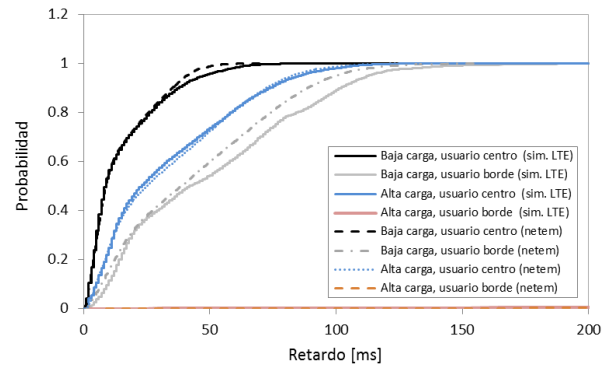


Fig. 12. Distribución acumulada de probabilidad del retardo total de paquete en el emulador y el simulador.

La tabla 4 resume los principales estadísticos de la transmisión de paquetes. Puede observarse otra vez el buen ajuste entre los valores obtenidos mediante el emulador y el simulador para la mayoría de los escenarios. En el caso (baja carga, usuario de centro), el error relativo en el retardo medio es inferior al 0.03%. En el caso (alta carga, usuario de centro), es del 4%, mientras que (baja carga, usuario de borde), el error relativo en el retardo medio es del 10%. Estos resultados son acordes a lo observado en la Fig. 12. En el caso (alta carga, usuario de borde), donde el sistema es inestable, la desviación en el retardo medio es considerable.

Tabla 4. Estadísticos de transmisión de paquetes.

Carga de red	Baja		Alta	
	Centro	Borde	Centro	Borde
Ubicación				
Retardo total medio (sim. LTE) [ms]	14.98	48.82	30.92	4543
Retardo total medio (NetEm) [ms]	14.19	43.71	33.06	21278
Jitter medio (sim. LTE) [ms]	2.09	4.42	3.86	9.86
Jitter medio (NetEm) [ms]	2.00	4.53	3.77	6.43

En la tabla 4, también se observa que el *jitter*, en todos los casos, tiene un valor del orden del retardo de paquete primero en cola. Como cabría esperar, la diferencia del retardo entre dos paquetes consecutivos de la misma ráfaga (cuadro de vídeo) se aproxima al retardo medio de transmisión primero en cola, tal como se deduce de la Fig. 7. Comparando los valores del simulador y el emulador, se deduce que el error relativo en el *jitter* está por debajo del 5% en todos los casos donde el sistema de cola es estable.

Tabla 5. Estadísticos de reproducción del vídeo.

Carga de red (porcentaje)	Baja (25%)		Alta (75%)	
	Centro	Borde	Centro	Borde
Ubicación				
Nº cuadros duplicados descompresor	16	12	46	585
Nº cuadros perdidos descompresor	3	0	0	6

En la Tabla 5 se presentan diversos estadísticos de rendimiento asociados a la decodificación del vídeo. Se observa cómo, en los tres casos donde el sistema es estable, el número de cuadros duplicados o perdidos, es muy inferior al obtenido cuando el sistema es inestable. Aun así, de las cifras de la tabla no se pueden extraer conclusiones relevantes sobre cuál de los tres casos es más favorable, dado que las cifras de cuadros perdidos/duplicados son muy pequeñas en cualquiera de los casos, lo que las hace dependientes de la realización específica de las variables aleatorias del experimento.

VI. CONCLUSIONES

En este trabajo se ha presentado una maqueta de sistema para la evaluación de la calidad de experiencia del servicio de *videostreaming* 3D en una red LTE. Las aplicaciones que conforman la maqueta son *VLC* como servidor y cliente de *videostreaming*, *NetEm* como emulador de red y *Wireshark* como herramienta de monitorización.

Para modelar el retardo de transmisión de forma precisa, se ha modificado el código fuente de *NetEm* para implementar un sistema de colas, donde el tiempo de respuesta del sistema es la suma del retardo de encolado y el retardo una vez que el paquete es primero en la cola. Esta última componente del retardo se obtiene a partir de las distribuciones de retardo de paquete obtenidas en un simulador LTE de nivel de sistema configurado para modelar escenarios realistas.

Los resultados han demostrado que, a diferencia del emulador original, la maqueta permite modelar de forma bastante precisa el retardo medio y el *jitter* bajo diferentes condiciones radio. Concretamente, en los casos donde el sistema es estable (es decir, cuando la tasa de transmisión de datos media supera a la tasa media de codificación del vídeo), el error relativo del retardo y el *jitter* respecto al simulador LTE es inferior al 10% y el 5%, respectivamente.

Se tiene previsto extender el presente estudio con la realización de una campaña de medidas de la calidad subjetiva experimentada cuando las secuencias se reproducen con un terminal móvil con pantalla 3D. Asimismo, se ampliará la maqueta con nuevos formatos de codificación de vídeo (p.ej., vídeo y profundidad) y métodos de medida de calidad objetiva de vídeo 3D (p.ej., PSNR, MS-SSIM).

AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad (proyecto TIN2012-36455) y la Junta de Andalucía (P10-TIC-6897). Asimismo, los autores quieren agradecer la colaboración de Almudena Sánchez Martínez y Francisco Blánquez Casado en la realización de los experimentos.

REFERENCIAS

- [1] I. Kulik, T. A. Trinh, "Investigation of Quality of Experience for 3D Streams in Gigabit Passive Optical Network", EUNICE 2011, LNCS 6955, pp. 157-168, 2011.
- [2] Z. Liu, et al., "Experimental Evaluation of H.264/Multiview Video Coding over IP Networks", ISSC 2011, Jun 2011.
- [3] C. Jean, et al., "QoE-based Packet Drop Control for 3D-Video Streaming over Wireless Networks", LANC'12, pp. 59-66, Oct 2012.
- [4] S. Singh, et al. "Video Capacity and QoE Enhancements over LTE", ICC 2012, pp. 7071 - 7076, 2012.
- [5] I. Palomo, M. Solera, M. Toril, J. Poncela-González, G. Gómez "Maqueta de sistema para la evaluación del servicio de vídeo streaming 3D sobre LTE", URSI 2014, Valencia (España), Sep. 2014.
- [6] G. Gómez, et al., "A Real-Time End-to-End Testbed for Evaluating the Performance of Multimedia Services", FMN 2009, LNCS 5630, pp. 212-217, 2009.
- [7] P. Uppu, S. Kadimpati, "QoE of Video Streaming over LTE Network", MS Thesis, Oct 2013.
- [8] <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>.
- [9] J. Sliwinski, A. Beben, P. Krawiec, "EmPath: Tool to Emulate Packet Transfer Characteristics in IPNetwork", Lecture Notes In Computer Science, Springer, nr 6003, 2010.
- [10] G. Gomez, D. Morales-Jimenez, J. J. Sanchez-Sanchez, J. Tomas Entrambasaguas, "A Next Generation Wireless Simulator based on MIMO-OFDM: LTE Case Study", EURASIP Journal on Wireless Communications and Networking, Vol. 2010, 2010.
- [11] <https://www.ffmpeg.org/>.

Design of smart services and routing protocols for VANETs in smart cities

Workshop QoS y QoE en Comunicación Multimedia

Ahmad Mohamad Mezher, Cristhian Iza Paredes, Luis Urquiza-Aguiar, Ángel Torres Moreira, Mónica Aguilar Igartua

Department of Network Engineering.

Universitat Politècnica de Catalunya (UPC)

C/ Jordi Girona 1-3, 08034 Barcelona, Spain.

[ahmad.mezher, cristhian.iza, luis.urquiza, maguilar]@entel.upc.edu, angel.torres.moreira@gmail.com

Abstract—In this paper, we introduce some proposals to tackle different problems present in mobile ad hoc networks (MANETs) and in vehicular ad hoc networks (VANETs). Firstly, we describe a multipath routing protocol to provide video-warning messages in MANETs using a game-theoretical approach. Secondly, we summarize a routing protocol for VANETs that uses four different metrics plus a proposal called Coherent, Automatic Address Resolution (CAAR). Thirdly, we summarize some works to disseminate emergency messages in realistic vehicular scenarios as a starting point to design our own proposal. All simulation results show the benefits of our proposals. Finally, we introduce an Android application developed to gather geolocation data of the users to analyse the mobility of citizens in the metropolitan area of Barcelona. The goal of this application is to improve the current transportation system in smart cities.

Keywords—Mobile ad hoc networks, Vehicular ad hoc networks, Smart cities, Adaptive multi-path routing protocols, Game theory, Video-streaming services, Multi-metric forwarding decisions, Geographic routing protocol, Address resolution (AR), Neighbour discovery (ND), Android applications

I. INTRODUCCIN

Smart city promising new services that involve vehicles, drivers, citizens and city infrastructures are about to emerge. Emergency prevention and response are key issues for smart cities to face any kind of sudden incident such as traffic accidents, traffic jams, public transport delays, *etc.* The prevention of accidents is one of the most important goals in smart cities, and nowadays Information and Communication Technologies (ICTs) and citizens play an essential role in this field. A good management of video-warning messages after accidents would lead to proper actions such as an immediate call to health car units and ambulances, warn other citizens around, change the frequency of traffic lights, change the schedule of buses, and so on. In these kinds of situations, MANETs, VANETs and citizens play an important role in the management of these smart cities. In [1], we have designed a game-theoretical multi-path routing protocol for MANETs to send video-warning messages in a smart city.

On the other hand, vehicular ad hoc networks (VANETs) [2] are an emerging area of wireless networking that facilitates ubiquitous connectivity among smart vehicles through vehicle-to-vehicle (V2V) communications and through vehicle-to-infrastructure (V2I) communications. This emerging technology field aims to improve the safety of passengers, alleviate the traffic flow, reduce pollution and enable in-vehicle entertainment applications for passengers.

To address these challenges in both safety and comfort-oriented applications, there is a need to develop new routing protocols specially designed for this kind of network, which will perform either under sparse or dense traffic conditions. In [3], we proposed a new routing protocol for VANETs which considers several metrics to select the best next forwarding node for each packet in each step towards its destination.

Besides, due to the very dynamic topology of VANETs, their routing protocols need continuous broadcasting of signaling messages which are used to keep track changes in the network infrastructure. On the other hand, an address resolution mechanism is traditionally required to set up a traffic flow when a new IP communication process starts. In [4], we show that this address resolution task can be performed by the network-topology updating process accomplished by VANET routing protocols. The new scheme to carry out the Address Resolution (AR) process in ad hoc networks was called Coherent, Automatic Address Resolution (CAAR).

Moreover, a general evaluation of dissemination multi hop protocols in VANETs is also done in this article. Finally, an Android Application able to obtain mobility data of the citizens in the metropolitan area of Barcelona is implemented. The rest of the paper is structured as follows. Section II introduces a game-theoretical routing scheme to send video-warning messages over MANETs. Section III presents a Multi-Metric Map aware routing protocol (MMMR) and a Coherent, Automatic Address Resolution (CAAR) for VANETs. In section IV, we study the performance evaluation of dissemination protocols for emergency messages in VANETs. Section V explains an android application named *Mobilitapp* that obtains mobility data of the citizens in the metropolitan area of Barcelona. Finally, conclusions are given in section VI.

II. GAME-THEORETICAL ROUTING SCHEME FOR VIDEO-STREAMING IN MANETs

A *game* can be described by listing the players participating in the game, a set of strategies for those players, and a specification of payoffs for each combination of strategies. Let S be a finite set of N players $1, \dots, N$. Each player i has a finite set of available actions A_i . Let $a_i \in A_i$ be each particular action chosen by player i . The action space, A , is the cartesian product of all A_i , *i.e.*, $A = A_1 \times A_2 \times \dots \times A_N$. An N -tuple action, a , is a point in the action space A . A *pure strategy*

provides a complete definition of how a player will play a game. In particular, it determines the move a player will make for any situation it could face. A *mixed strategy* of player i , α_i , is an assignment of a probability, $p_i \in P = [0, 1]$, to each pure strategy. This allows a player to randomly select among the set of pure strategies. Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$ be the mixed strategy profile, then the probability that a particular N -tuple action, $a = (a_1, a_2, \dots, a_N)$, will occur, $p(a)$, is formed from the product of the probabilities assigned to a by α . Let u_i be the *utility function* of player i in the strategic form game occurring in each stage. The utility function is a mathematical description of preferences that maps the action space to a set of real numbers. A utility function for a given player assigns a real number for every possible outcome of the game, so that a higher number implies that the outcome is more preferred.

$$u_i : A \rightarrow R \quad (1)$$

$U_i(\alpha)$ is the expected utility for player i for the mixed strategy profile α , and has the following expression:

$$U_i(\alpha) \equiv \sum_{a \in A} p_a \cdot u_i(a) \quad (2)$$

A strategic game G can be expressed using three primary components: the set of players S , the action space A , and the set of individual utility functions for player i , u_i .

$$G = (S, A, u_i) \quad (3)$$

A mixed strategies extension to G is given by the next expression, where $\Delta(A_i)$ is the set of all probability distributions over A_i and U_i is the set of all expected utilities to i .

$$G' = (S, \Delta(A_i), U_i) \quad (4)$$

A *best response* is a strategy which produces the most favorable outcome for a player, taking other players' strategies as given. A *Nash Equilibrium* (NE) [5] is a solution in which each player plays a best response to the strategies of other players. Each player is assumed to know the strategies of the other players, and no player has incentive to unilaterally change their current strategy while the other players keep theirs unchanged. Players are in equilibrium if a change in strategies by any one of them would lead that player to earn less than if they remained with their current strategy. It is a mathematical fact that every mixed extension of a strategic game has at least one *mixed strategy Nash equilibrium* [5].

To define best responses more generally, we need a notation for the set of strategies used by all players other than player i , named α_{-i} :

$$\alpha_{-i} = (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_N) \quad (5)$$

Strategy α_i^* is a best response for player i to the strategies of all players except i , α_{-i}^* , if:

$$U_i(\alpha_i^*, \alpha_{-i}^*) \geq U_i(\alpha_i, \alpha_{-i}^*), \quad \forall \alpha_i \in \Delta(A_i) \quad (6)$$

This means that if α_i^* is a best response for player i to the assumed set of strategies α_{-i}^* played by the other $N-1$ players, then it must give player i a payoff at least as large as the player would get if they placed any other strategy α_i from

their set of allowed strategies. Equivalently, a best response (BR) correspondence to player i is given by :

$$\alpha_i^* \in BR_i(\alpha_{-i}) = \operatorname{argmax}_{\alpha_i \in \Delta(A_i)} U_i(\alpha_i, \alpha_{-i}) \quad (7)$$

Let us remark that $\operatorname{argmax}_x F(x)$ is the value of x for which $F(x)$ has the largest value. A joint strategy $\alpha^* = (\alpha_1^*, \dots, \alpha_N^*)$ is a NE if, for each player i , α_i^* is a best response to α_{-i}^* .

Remember that every finite game (finite number of players, each of which with a finite set of strategies) has Nash equilibria in either mixed or pure strategies [6].

A. The Benefit of Using a Particular Path to Transmit the I+P Video Frames

Before defining the player's utility of the game, we will define a parameter that evaluates the benefit of using a particular path. As we have mentioned before, we assume that we always have at least two available paths (best and medium-quality paths) to send packets. Each path will have its own benefit as we describe next.

Definition: The benefit of a path k is:

$$\beta_k = \begin{cases} \phi_k, & \text{if the path fulfills the QoS requirements} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\phi_k \in \mathbb{R}^*$.

Let us assign $\phi_k = \phi_b$ as the benefit for the best path and $\phi_k = \phi_m$ as the benefit for the medium-quality path, where ϕ_b and $\phi_m \in \mathbb{R}^*$. Later, we will relate ϕ_b and ϕ_m with the subjective QoS experienced by the users regarding the video frames received from each path. Strategy α_i is defined as follows:

$$\alpha_i = \begin{cases} \text{Transmit using the best path.} \\ \text{Transmit using the medium-quality path.} \end{cases} \quad (9)$$

Probability p is the probability of sending (I+P) frames through the best path and probability $(1-p)$ is the probability of sending those frames through the medium-quality path.

B. Design of the Utility Function

The utility function U_i designed for our game-theoretical routing protocol aims at achieving two goals: Minimizing the rate of (I+P) frames lost and minimizing the delay jitter, since these two parameters are most significant in video-streaming services. An initial short delay can be easily bearable by the user. Part of that delay would be produced after gathering frames in the reception buffer to reorder frames before the decoding process. The proposed utility function for player i is the following:

$$U_i = \underbrace{\left(\frac{n_{rb,i} - n_{sb,i}}{n_{sb,i}} \right) \cdot \phi_{b,i} \cdot \frac{p_i^2}{J_{b,i}}}_{\text{Best path}} + \underbrace{\left(\frac{n_{rm,i} - n_{sm,i}}{n_{sm,i}} \right) \cdot \phi_{m,i} \cdot \frac{(1-p_i)^2}{J_{m,i}}}_{\text{Medium path}} \quad (10)$$

where, $i = 1, 2, 3, \dots, N$ is a generic player, being N the number of players; p_i is the player's i probability of sending the (I+P) frames through his/her best path; $\phi_{b,i}$ is the player's i benefit for the best path; $\phi_{m,i}$ is the player's i benefit for the medium-quality path; $n_{sb,i}$ is the number of (I+P) frames sent by player i through the best path; $n_{rb,i}$ is the number of (I+P) frames received from player i through the best path; $n_{sm,i}$ is the number of (I+P) frames sent by player i through the medium-quality path; $n_{rm,i}$ is the number of (I+P) frames received from player i through the medium-quality path; $J_{b,i}$ is the delay jitter of player i through the best path and $J_{m,i}$ is the delay jitter of player i through the medium-quality path.

To conclude with, Algorithm 1 summarizes the methodology to compute the best response probability p_i^* for player i to send his/her I+P frames through the best path. For further information, please check [1].

Algorithm 1 Calculation of p_i^* , the best response probability for player i that maximizes his/her utility function U_i .

Require: Obtain updated QoS values from the periodically received RTCP packets.

- 1: Obtain the values of $(MOS_b, MOS_m, J_b, J_m, \hat{n}_b, \hat{n}_m)$
- 2: Compute the $k_{b/m}$ parameter designed that fulfills the requirements.

$$k_{b/m} = \frac{\pm \sqrt{\frac{MOS_b}{J_b} \cdot \frac{MOS_m}{J_m} \cdot (\hat{n}_b + \hat{n}_m - 2)} - \frac{MOS_m}{J_m}}{2 \cdot \frac{\partial p_i^*}{\partial k_{b/m}} \frac{MOS_b}{J_b}}$$

- 3: Calculate the probability p_i^* to send I+P frames through the best path, that maximizes the utility function U_i

$$p_i^*(k_{b/m}) = \frac{\left(\frac{k_{b/m} \cdot MOS_b}{J_b}\right) \cdot \hat{n}_b + \left(\frac{MOS_m}{J_m}\right) \cdot (2 - \hat{n}_m)}{2 \cdot \left(\frac{k_{b/m} \cdot MOS_b}{J_b} + \frac{MOS_m}{J_m}\right)}$$

C. A Case Study in a Smart City

We focus our research work on two realistic smart city scenarios. A mobile adhoc network (MANET) and a vehicular ad hoc network (VANET), where emergency prevention and response are key issues. In the two scenarios under consideration, we assume that in a given moment an accident happened. Most of the citizens nowadays carry mobile phones or tablets. In the MANET/VANET scenario, we assume that a smart citizen/driver witnesses the situation, makes a short video-warning message (the driver will just push a button that will make a small exterior car-mounted camera shoot the video) about the accident and sends it through the MANET/VANET to the nearest emergency unit (e.g., police, ambulances, hospitals). Authorities will respond upon receiving the video and will take proper actions. This way, with a video-warning message the emergency can be evaluated much better than with a simple text. It would be easier to ensure an accurate interpretation of the situation and the accident could be treated with the adequate level of seriousness. The smart citizen/driver sends a multimedia message which includes different information regarding the incident, e.g., the GPS location, a voice message and a short video of the incident. A suitable kind of smart-112 application in the citizen's mobile/vehicle sends the multimedia message to the smart-112 emergency center, who manages the proper actions for that incident. For instance,

ambulances and paramedical will be sent there, traffic lights will turn to red around the accident, a green wave of traffic lights will help the ambulances get there sooner, the nearest hospital is warned, the doctors wait for the injuries, etc. Our purpose in this work is to design a game-theoretical multipath routing protocol suitable to transmit those video-warning messages over MANETs/VANETs in this kind of smart city scenarios.

D. Performance Evaluation in a VANET Scenario

In this subsection, we introduce a brief performance evaluation of our proposal in a VANET scenario. Video flows are transmitted from two vehicles to two access points AP1 and AP2 (see Figure 1), respectively. AP1 is the Ana Torres Institute (a surgery clinic) and AP2 is the Hospital Clinic of Barcelona, which represent two emergency units where vehicles will send their multimedia warning messages upon the event of a traffic accident. Each vehicle will send its multimedia warning message to the closest AP. We carried out ten simulations per point using the NS2 [7]. Figure 2 shows the results with confidence intervals (CI) of 90 percent. In the simulations, we used a real city area obtained from the example district of Barcelona (see Figure 1). In order to simulate a realistic scenario, the CityMob for Roadmaps (C4R) [8] simulator was used to obtain the mobility model. Besides, C4R imports maps directly from OpenStreetMap [9] and generates NS2 compatible files to specify the mobility model for the vehicles through the city along the whole simulation. Area of the scenario is 850×580 m²; simulation time is 250 s and number of nodes are 50. Detailed simulation settings can be seen in Table 6 from [1].

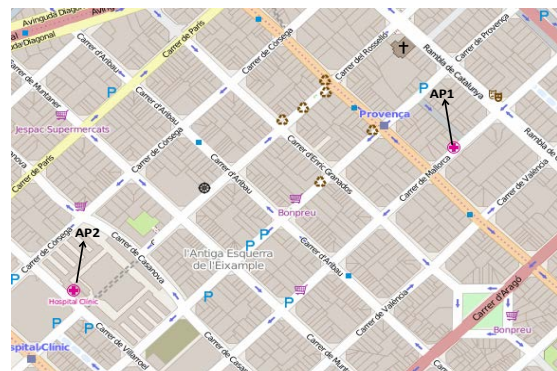


Fig. 1. Simulation scenario of Barcelona with 2 users, 50 vehicles and 2 emergency units: AP1 (Ana Torres surgery clinic) and AP2 (Hospital Clinic of Barcelona).

Figure 2(a) shows the average percentage of packet losses for $N = 2$ users when using the game-theoretical scheme against the case of non using it (*Non game* option). We can see how including the game-theoretical routing scheme, the average video packet losses are reduced from 26% to 11%. This decrement of the average packet losses is due to the optimal selection of paths based on a probability value (*i.e.*, p^*) that optimally balances the load among the two paths at stake (*i.e.*, the best and the medium-quality paths) depending

on the quality of the paths. Figure 2(b) depicts the average end-to-end packet delay. We can see that the case including the *Game* scheme improves the delay compared to the *Non Game* case. It is reduced from 0.09 s to 0.04 s. This first performance evaluation of our game-theoretical multi-path routing approach applied in VANETs shows clear benefits. Based on these incipient results, in a next work we will further develop a multi-hop geographical routing protocol for VANETs based on a game-theoretical approach to send video warning messages in a smart city. An extensive performance evaluation in the MANET scenario can be seen in [1] where we show the improvement of our game-theoretical approach.

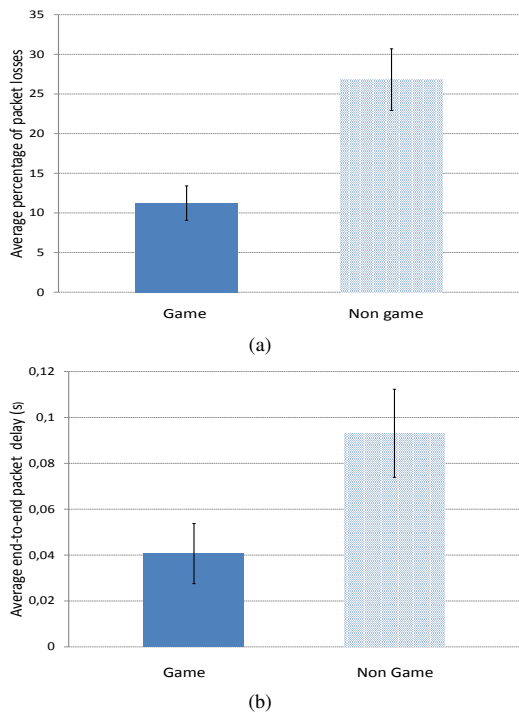


Fig. 2. Simulation results for $N = 2$ users. (a) Average percentage of packet losses; (b) Average end-to-end packet delay.

III. PERFORMANCE IMPROVEMENTS OF VANET ROUTING PROTOCOLS ORIENTED TO TRAFFIC EFFICIENCY

All the traffic efficiency-oriented applications require a continuous monitoring phase of the streets and city conditions. Vehicles are the most suitable collectors of this information so they can be in charge of feeding the monitoring centers. This kind of applications does not have important delay constraints, so the transport of its traffic is suitable for delay-tolerant protocols which deal with variable and high delay, since the data is still useful. This section presents some works that aim to improve the end-to-end performance of unicast routing protocol using during the phase of collecting data only.

A. Multi-Metric Map aware Routing protocol

First we propose a geographical routing protocol Multi-Metric Map aware Routing (MMMR) [3] designed for delay tolerant applications that seeks to improve Greedy Perimeter Stateless Routing (GPSR) [10] performance. GPSR was specially designed for VANETs and based its forwarding decision in the geographical distance between nodes. By default, the

neighbor geographically closest to the packet's destination is chosen as the packet's next hop, this behavior is named as *greedy forwarding*. When there is not a closest neighbor to destination, GPSR uses the right-hand rule to route packets around voids. The sequence of edges traversed by the right-hand rule is called *perimeter*.

MMMR improves GPSR because uses more information to get a more accurate position estimation, thus being more restrictive in selecting neighbors (e.g., avoiding nodes with low power reception signal). In addition, MMMR stores packets in a buffer (carry packets with it) instead of implementing perimeter mode, which is considered an inefficient recovery mode.

MMMR considers four metrics instead of only one as GPSR to select the next hop among its neighbors. These four metrics evaluate the distance to destination, the vehicle trajectory, the vehicle density, and the available bandwidth. The value of metric is computed through 11

$$u_{n,i} = e^{-f_i(x_{n,i})} \quad i = 1..4, \quad n \in \text{Nodes} \quad (11)$$

Where $u_{n,i}$ is the value of the metric i for the node n and it is between 0 and 1. This is because f_i that evaluates the components employed by the corresponding metric returns values on $[0, +\infty)$. Notice that f_i depends on the metric that is evaluated.

A node evaluates and assigns a total multi-metric qualification to each neighbor when it will route a packet. MMMR assigns equal weights (w_1, w_2, w_3, w_4) to each metric (u_1, u_2, u_3, u_4), in the qualification of each neighbor. The global geometric metric is computed with eq. (12). The neighbor with the highest combined score is chosen as next hop.

$$\bar{u} = u_1^{w_1} \cdot u_2^{w_2} \cdot u_3^{w_3} \cdot u_4^{w_4} \quad (12)$$

MMMR was evaluated under dense scenario. Results show that MMMR improves significantly, in terms of packet losses the other protocols in a dense scenario (See Fig. 3). This fact confirms the benefits of considering more metrics than only using the distance.

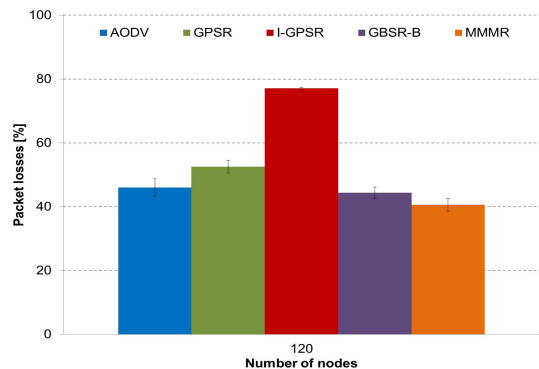


Fig. 3. Percentage of losses in dense scenario. (CI 99%).

B. Coherent, Automatic Address Resolution (CAAR)

Due to the very dynamic topology of VANETs, their routing protocols need continuous broadcasting of signaling messages which are used to keep track changes in the network infrastructure. On the other hand, an address resolution mechanism

is traditionally required to set up a traffic flow when a new IP communication process starts.

We show in [4] that this address resolution task can be performed by the network-topology updating process accomplished by VANET routing protocols. We called Coherent, Automatic Address Resolution (CAAR) to this new scheme to carry out the Address Resolution (AR) process in Ad Hoc networks. CAAR is really straightforward. Every node copies its MAC address, available for instance from its MIB (Management Information Base), into the signalling routing messages as a new field of them. This action will slightly increase the length of those signalling messages in 6 bytes with the purpose of not augmenting the MAC layer complexity or cross layering cooperation between MAC and routing layers. When a routing signaling packet arrives to a node, it is processed by the routing daemon which extracts the source IP and MAC addresses from the IP header and the signaling message, respectively. We do not need to use timers, since the match is directly done when an IP Packet encapsulated in a routing message is received. Such IP packets contain both IP and MAC addresses at the same communication layer. Next, the pair of IP and MAC addresses is sent to the Address Resolution process, which adds or updates an entry to the AR table. We also propose an improved AR, called AR+, that makes AR more compatible with characteristics of a VANET. AR+ modifies the behavior of AR learning, match its buffer size with routing buffer. Moreover, AR+ flush its AR table according to the timers set in the routing layer, avoiding unnecessary AR transactions.

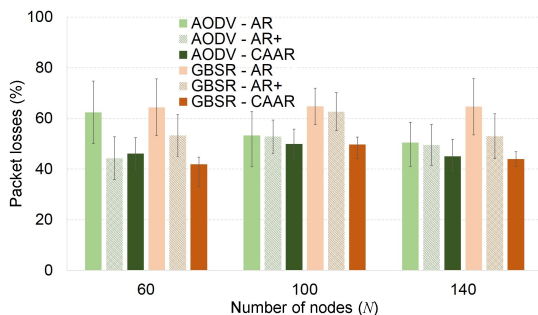


Fig. 4. Percentage of packet losses for AODV and GBSR using the three Address Resolution schemes. (CI 95%).

We evaluated the behavior of CAAR and compared it to traditional AR and AR+, in a realistic VANET scenario for three different vehicle densities. We noticed that CAAR (see Fig. 4), when implemented over geographical protocols like GBSR, has an improved performance in terms of packet losses compared to the other schemes we evaluated. However, when implemented over classical topological protocols like AODV, CAAR only shows significant improvements in low-density scenarios.

C. MAX - MIN based Buffer Allocation for VANETs

The MAC specification of the IEEE 802.11p standard uses the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) to provide four different Quality of Service (QoS) levels. Many research works focused on the optimization of the EDCA process to maximize the channel occupancy,

minimize the delay or guarantee reliability in the packet transmission. However, in [11] we focus our attention on the queuing process, which stores packets until an AC can access the medium. We proposed a buffer allocation mechanism for the four different access categories based on the aforementioned MAX-MIN principle, but modifying its behavior when there is not enough buffer room. Basically, the key ideas of our proposal are:

- (1) the joint of the buffer of the four access categories in a unique equivalent common buffer.
- (2) the buffer rooms are allocated according to incoming demand across the ACs
- (3) when the shared buffer does not have available room to serve an AC, our allocation process releases memory by dropping packets starting from the lowest to the highest AC priority. The aim is to follow a fairness criteria but keeping a precedence order in the queues during this operation. We named this proposal Prioritized MAX-MIN Buffer Allocation (PMMBA). We tested our mechanism with fixed guaranteed equal queue size PMMBA- F_Q and PMMBA- D_Q where the size of the queue depends on the losses probability associated with the priority of the traffic and computed according to the formulas of [12].

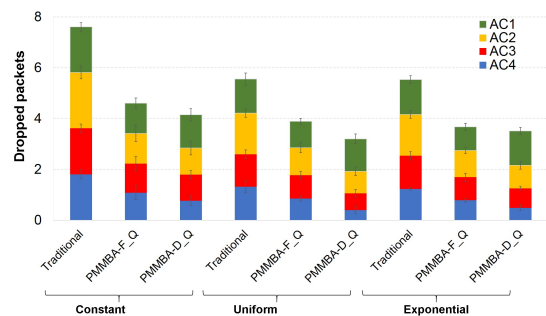


Fig. 5. Average dropped packets per node with PMMBA

Our results in figure 5 shows the average dropped packets per node in the AC queues during the simulations. It allow us to see that even in low rate experiments, there are buffer overflows. The overflows occur when a big enough group of nodes select the same next forwarding node, because it is the closest one. This situation clearly entails a buffer saturation. Last but not least, Fig. 5 gives the chance to realize, for PMMBA- D_Q , the precedence order in the guaranteed size among AC queues. Notice that the drops in AC₄ are lower in PMMBA- D_Q than the other techniques thanks to the lowest P_L set and the precedence order of our algorithm.

IV. PERFORMANCE EVALUATION OF DISSEMINATION PROTOCOLS FOR EMERGENCY MESSAGES IN VEHICULAR AD-HOC NETWORKS

In this section, we focuses on the evaluation of dissemination multi hop protocols in VANETs. We have selected some of the most representative dissemination protocols proposed in the current literature to evaluate them. In our evaluation, we have used realistic scenarios which take into account the most influential factors on the performance evaluation. Comparative results of the assessment based on evaluation metrics are presented.

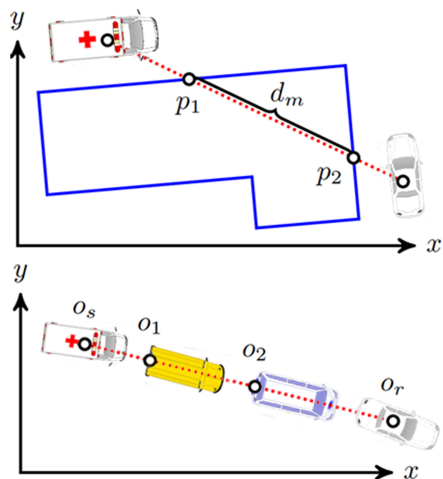


Fig. 6. Calculation of line of sight intersection points with building and vehicles [17]

A. Fair evaluation in realistic scenarios

According to the authors in [13], five factors have a strong influence on the quality of performance evaluation of protocols in vehicular environments, which are: a realistic mobility pattern of vehicles in the simulation, a realistic scenario, propagation models of the radio signal, appropriate evaluation metrics and realistic human driver behavior.

To fulfill with these factors, we have selected VEINS [14] to carry out our fair simulations. VEINS has important features such as simulation framework for vehicular networks, on-line reconfiguration and re-routing of vehicles, fully detailed models of the IEEE 802.11p standard, IEEE 1609.4 WAVE, support for realistic maps and realistic traffic. In order to achieve the highest level of realism possible, we use real maps extracted from OpenStreetMap [9].

The biggest challenge facing VANET networks in urban scenarios is the presence of buildings and other artificial structures that affect wireless connectivity (see Fig. 6). Shadows in radio communication produced by buildings is an important factor for road safety critical applications. In [15], authors presented a computationally affordable simulation model for obstacles with IEEE 802.11p in urban environments. It is an empirical model based on measurements of the real world, because it only considers the line of sight between transmitter and receiver. This model of obstacles is integrated with the framework of VEINS. The impact of shadows on the radio communication caused by vehicles was evaluated in [16]. The aim of the study was to show that the impact of the blockage of vehicles on the radiocommunications is not negligible. Based on this model, we have added an additional calculation of the attenuation due to vehicles in the framework of VEINS. To compute the impact of moving vehicles on the power loss, we employ a similar technique to those presented in [16] and [17].

On the other hand, researchers only present the metrics which are advantageous to the proposed protocols, while other metrics are ignored. Therefore, besides using conventional metrics, we assess the performance of protocols based on the unified metric presented in [18]. This metric is called Dissemination Efficiency (DE) and it is a measure that includes

three separate domains (frequency, time and space). The DE metric is defined as:

$$DE = \frac{\text{Propagation Distance} \times \text{Success Rate}}{\text{Propagation Time} \times \text{Redundancy Rate}} \quad (13)$$

Intuitively, DE measures how far an information packet can propagate through the network per unit of time and per amount of overhead generated.

B. Simulation setup

To carry out the performance of the analysed dissemination schemes, we run 10 repetitions for each different simulation experiment for statistical confidence and we show the 95% confidence intervals in all figures. The packet error and Medium Access Control (MAC) layer models adopted are based on the IEEE 802.11p. For eliminating effects caused by switching channel between the Control Channel (CCH) and the Service Channel (SCH), we changed the model to use only the CCH. Bandwidth is equal to 10 MHz; transmission range is 230m; transmission power is 20 mW; sensitivity is equal to -94 dBm; (CWmin, CWmax)= (15,1023); AIFSN equal to 2 and the Bit rate is equal to 6Mbit/s.

For all simulation scenarios, the data message size is 2312 bytes, i.e. the maximum allowed by the 802.11p standard. This allows us to evaluate the protocols in the worst-case scenario in terms of medium occupation caused by the transmission of messages. Beacons are sent at the frequency of 1 Hz. This is usually the highest frequency expected to be used for the transmission of beacons which gives the worst-case scenario in terms of freshness of the one-hop neighborhood information. In order to study realistic vehicle-caused radio shadowing, we used a typical mix of different vehicles (90% cars and 10% trucks). Trucks can attenuate or even block the signal generated from cars. All vehicles are moving according to the SUMO standard Krauss driver model.

C. Scenario Description

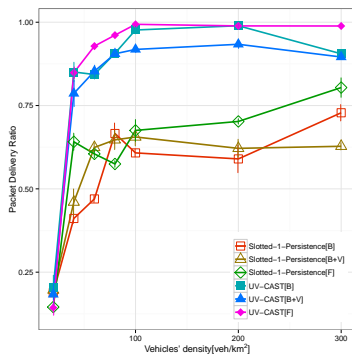
We focus on the immediate consequences of an accident. The crashed vehicle starts to generate and transmit an emergency message after the collision to inform neighboring vehicles as quickly as possible in a distributed way. In the simulations, we used a real city area obtained from Barcelona, Spain as our urban scenario. This segment has an area of 1.5 x 2km² and was retrieved from OpenStreetMaps [9]. A vehicle positioned approximately at the center of the network is responsible for generating a single message to be disseminated in a TTL of 90s.

For all the experiments, the results for the three types of shading models are presented: Clearances (F), shadow Building (B), shadow Buildings + Shadow vehicles (B+V).

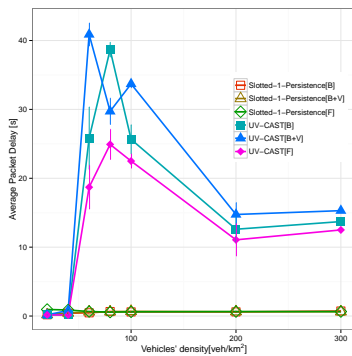
D. Simulation Results

In this section, we present some representative simulation results. We select two state-of-the-art protocols for comparison, called: UV-CAST [19] and Slotted 1-Persistence [20].

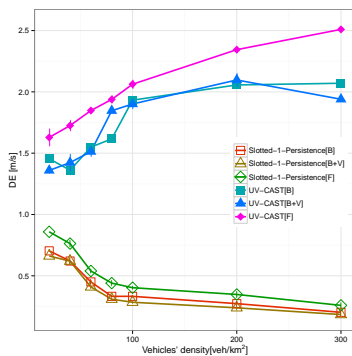
In a first set of experiments, we evaluated the performance of the packet delivery ratio. Fig. 7(a) shows the packet delivery rate for three radio shadowing models. In the case of free space (F), UV-CAST achieves successful delivery rates close to 100% above 100 vehicles/km². The protocol uses only a



(a) Packet delivery ratio



(b) Average packet delay



(c) Dissemination Efficiency

Fig. 7. All figures present 95% confidence intervals for different network densities in an urban scenario.

subset of vehicles the task of opportunistic forwarding (store-carry-forward). Thus, uninformed vehicles that do not find a vehicle from this subset will not receive the disseminated message. It is clear the poor performance of Slotted-1-Persistence for low traffic density (3 inferior lines in Fig. 7(a)). This result is expected, because this protocol was designed for high density. However, Slotted-1-Persistence only reaches delivery rates that do not exceed 80%. As this protocol does not have a mechanism for opportunistic forwarding, dynamic topology networks in VANET causes temporary disconnections, interrupting the dissemination and compromising the delivery of the message. There is a difference between packet delivery rate in free space (F) and delivery rate with shadows of buildings (B). This difference is even more visible if building

and high vehicles (B+V) are considered. We believe that the ability of the dissemination protocols conceals a greater impact of the shadows from buildings and high vehicles.

In a second step, we investigate the performance of latency. Fig. 7(b) shows the average packet delay for three radio shadowing models. In the case free space (F), when the traffic density is low, UV-CAST has a very low delay (2-5 s). This is due to the fact that the protocol fails to deliver the message to almost all vehicles of the region of interest as it was shown in Fig. 7(a). On the other side, the high latency value in UV-CAST in scenarios with high vehicular density is due to the excessive number of broadcast messages. These messages increase contention and packet loss, and therefore delay in message dissemination. Similarly, the results presented in Fig. 7(b) confirm the impact that buildings and high vehicles have as obstacles in the line of sight in UV-CAST protocol. There is a difference between average packet delay in free space (F) and delivery rate with shadows of buildings (B). This difference is more pronounced if building and high vehicles (B+V) are considered. If the scenario is densely congested, the beacon messages have shorter range, i.e. the amount of vehicles that is able to receive beacons is affected by radio signal shadowing. It is important to say that the low average packet delay (2-5 s) of Slotted-1-Persistence is because it does not perform the task of store carry forward. Its performance is limited to direct relay through the dissemination of multi hop, which can be done very quickly.

Finally, in Fig. 7(c) the average values for the metric Dissemination Efficiency (DE) are presented for three radio shadowing models. As it can be seen in the case free space (F), DE values in the Slotted-1-Persistence protocol (3 lines at the bottom at Fig. 7(c)) decrease in higher traffic density. Although Slotted-1-Persistence was designed for high densities, the short propagation range and its high rate of packet losses do not allow to get a better performance. In contrast, as can be observed that in the case of UV-CAST (3 lines at the top of the Fig. 7(c)), DE values have a tendency to grow at high density. Undoubtedly, opportunistic forwarding mechanism allows to achieve a propagation distance greater and a higher delivery rate in comparison to Slotted-1-Persistence, therefore, a better performance in the dissemination. This metric confirms that the dissemination is affected by the patterns of shadows (B and B+V). We conclude that Dissemination Efficiency is a metric that allows a clear view of the performance of a dissemination protocol.

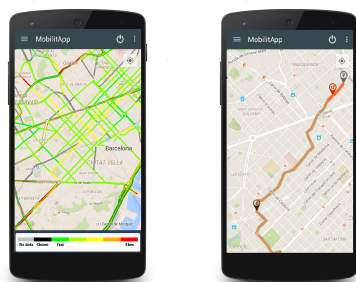
V. MOBILITAPP

In this section, we will introduce an android application called MobilitAapp (<http://mobilitapp.noip.me>) to obtain mobility data of the citizens in the metropolitan area of Barcelona. This study is part of the INRISCO (INcident monitoRing In Smart Communities) Spanish project. Our application is designed specifically to help the ATM (*Autoritat del Transport Metropolità*) by providing mobility data. Our implementation obtains GPS positions of citizens as well as their current activity (bus, metro, on foot...). At the end of the day, information data is processed and sent to a server where the information will be stored. With this information, we can

obtain the mobility patterns of citizens that could help the ATM to improve the current transportation service.

One of the main features added was the possibility to differentiate between metro, bus, train and tram when a *vehicle activity* is detected. Users are able to visualize in a map all their locations history that are stored in a daily base. In the last version of our application, more features were added, such as:

- Traffic state real-time information (Barcelona OpenData)
- Traffic incidences (*Dirección General de Tráfico*)
- Web application to analyze and filter the collected information about mobility



(a) Real-Time traffic state (b) Location History

Fig. 8. Features and graphical interfaces of MobilitApp.

VI. CONCLUSIONS

Simulation results in a VANET scenario using our game-theoretical routing protocol to send video-warning messages in a smart city outperform the results compared to the case of non using it, in terms of packets losses and delay. In addition, MMR improves all the protocols evaluated in terms of packets losses and throughput, due to the new way of selecting the next forwarding node. Also, we noticed that CAAR, when implemented over geographical protocols like GBSR, has an improved performance in terms of packet losses compared to the other schemes we evaluated. Moreover, after studying several proposals for dissemination of emergency messages in realistic urban scenarios and due to frequent topology changes and dynamic traffic density that VANETs present, new protocols should be designed taking into account the special characteristics of realistic vehicular networks. Besides, our android application introduced in this paper could help the City Halls to improve their public transportation systems.

ACKNOWLEDGEMENT

This work was partly supported by the Spanish Government through projects TEC2013-47665-C4-1-R (EMergency Response In Smart Communities, EMRISCO), TEC2014-54435-C4-1-R (INcident monitoRing In Smart Communities, QoS and Privacy, INRISCO) and AGAUR Information Security Group (ISG) project - 2014 SGR 1504. Ahmad Mezher is the recipient of a FI-AGAUR grant, from the Government of Catalonia. Cristian Iza and Luis Urquiza are recipients of a grant from Secretaria Nacional de Educación Superior, Ciencia y Tecnología SENESCYT.

REFERENCES

- [1] A. Mezher, M. A. Igartua, L. J. de la Cruz Llopis, E. P. Segarra, C. Tripp-Barba, L. Urquiza-Aguilar, J. Forné, and E. S. Gargallo, "A multi-user game-theoretical multipath routing protocol to send video-warning messages over mobile ad hoc networks," *Sensors Networks*, vol. 15, pp. 9039–9077, 2015.
- [2] H. Hartenstein and K. Laberteaux, *VANET Vehicular Applications and Inter-Networking Technologies*. Intelligent Transport Systems, 2009.
- [3] C. Tripp-Barba, L. Urquiza-Aguilar, M. Aguilar Igartua, D. Rebollo-Monedero, L. J. de la Cruz Llopis, A. M. Mezher, and J. A. Aguilar-Calderón, "A multimetric, map-aware routing protocol for VANETs in urban areas," *Sensors (Basel, Switzerland)*, vol. 14, no. 2, pp. 2199–224, Jan. 2014.
- [4] L. Urquiza-Aguilar, C. Tripp-Barba, D. Rebollo-Monedero, A. M. Mezher, M. Aguilar-Igartua, and J. Forné, "Coherent, Automatic Address Resolution for Vehicular Ad Hoc Networks," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. (in press), pp. 1–18, 2015.
- [5] J. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, pp. 286–295, 1951.
- [6] P. Dutta, "Strategies and games: Theory and practice," *MIT Press*, 2001.
- [7] "The network simulator, ns-2." [Online]. Available: <http://nslam.isi.edu/nslam/>
- [8] M. Fogue, P. Garrido, F. Martínez, J. Cano, C. Calafate, and P. Manzoni, "Realistic simulation framework for vehicular networks," in *Proc. 5th International ICST Conference on Simulation Tools and Techniques Simutools'12*, March 2012, pp. 37–46.
- [9] "Open street maps." [Online]. Available: <http://www.openstreetmap.org/>
- [10] B. Karp and H. T. Kung, "GPSR Greedy perimeter stateless routing for wireless networks," in *6th annual international conference on Mobile computing and networking*. New York, USA: ACM Press, 2000, pp. 243–254.
- [11] L. Urquiza-Aguilar, A. Vazquez-Rodas, C. Tripp-Barba, M. A. Igartua, L. J. de la Cruz Llopis, and E. S. Gargallo, "MAX-MIN based buffer allocation for VANETs," in *Wireless Vehicular Communications (WiVeC), 2014 IEEE 6th International Symposium on*, Sep. 2014, pp. 1–5.
- [12] A. Vázquez-Rodas, L. J. de la Cruz Llopis, M. Aguilar Igartua, and E. Sanvicente Gargallo, "Dynamic buffer sizing for wireless devices via maximum entropy," *Computer Communications*, Mar. 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0140366414000899>
- [13] C. Sommer and F. Dressler, *Vehicular Networking*. Cambridge University Press, November 2014.
- [14] "Veins, vehicular network simulations," <http://veins.car2x.org>, accessed: 2015-03-22.
- [15] C. Sommer, D. Eckhoff, R. German, and F. Dressler, "A computationally inexpensive empirical model of IEEE 802.11p radio shadowing in urban environments," in *Wireless On-Demand Network Systems and Services (WONS), 2011 Eighth International Conference on*, 2011, pp. 84–90. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5720204>
- [16] M. Boban, T. Vinhoza, M. Ferreira, J. Barros, and O. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks," vol. 29, no. 1, pp. 15–28, 2011. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5678778>
- [17] C. Sommer, S. Joerer, M. Segata, O. Tonguz, R. Lo Cigno, and F. Dressler, "How shadowing hurts vehicular communications and how dynamic beaconing can help," to be published, early Access. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6920079>
- [18] S. Panichpapiboon and W. Pattara-Atikom, "A review of information dissemination protocols for vehicular ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 3, pp. 784–798, 2012. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5989903>
- [19] W. Viriyasitavat, O. Tonguz, and F. Bai, "Uv-cast: an urban vehicular broadcast protocol," vol. 49, no. 11, pp. 116–124, 2011. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6069718>
- [20] N. Wisitpongphan, O. Tonguz, J. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," vol. 14, no. 6, pp. 84–94, 2007. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4407231>

QoE en redes adhoc, descarga adaptativa de contenidos y vídeo 3D

Pau Arce, Ismael de Fez, Francisco Fraile, Santiago González, Paola Guzmán, Juan Carlos Guerri
Grupo Comunicaciones Multimedia, iTEAM (Instituto de Telecomunicaciones y Aplicaciones Multimedia)
Universitat Politècnica de València
Camino de Vera, s/n
paarvi,isdefez,ffraile,sangonma,paoguzc1,jcguerri@iteam.upv.es

Resumen- El artículo resume algunas de las líneas de trabajo del grupo de Comunicaciones Multimedia del iTEAM (UPV) en relación a la QoE. Entre las diferentes propuestas, se describe el efecto de un algoritmo basado en nodos estratégicos para maximizar la QoE (tiempo de vida y PSNR) del servicio de videostreaming en redes Ad Hoc. En segundo lugar, se presenta una arquitectura híbrida multicast-unicast, basada en FLUTE y DASH para la descarga de contenidos (incluido vídeo). Dicha arquitectura aprovecha las características de ambos protocolos para maximizar la eficiencia del ancho de banda en la descarga de los contenidos. Por último, se analiza la QoE desde el punto de vista de la codificación de vídeo para contenidos 3D, considerando tanto versiones de codificación simétrica como asimétricas.

Palabras Clave- QoE, Videostreaming, Ad Hoc, DASH, FLUTE, 3D.

I. INTRODUCCIÓN

La temática relacionada con la Calidad de Experiencia (QoE, *Quality of Experience*) percibida por los usuarios sobre multitud de servicios (como por ejemplo, Videostreaming, IPTV, Navegación Web, etc.) se ha convertido en un tema de investigación con numerosas contribuciones en los últimos años. El concepto de QoE centrado en el usuario, en realidad complementa el concepto de la Calidad de Servicio (QoS, *Quality of Service*), centrado en aspectos técnicos (tasa de pérdidas, retardo, jitter, throughput, tiempo de zapping, etc.) de las redes que dan soporte a dichos servicios. Recientemente se han publicado unos artículos donde se revisan de forma muy exhaustiva los métodos de evaluación tanto objetivos como subjetivos relacionados con el servicio de videostreaming [1],[2]. El punto de partida en ambos casos es la importancia que el tráfico de vídeo tiene actualmente en Internet y que tendrá en el futuro. Según las estadísticas disponibles y previsiones, el tráfico basado en streaming de vídeo constituirá el 69% de tráfico de Internet en 2017 [3], siendo Youtube, Netflix y Hulu los proveedores de servicios de videostreaming más populares. Desde el punto de vista del modo de transporte del tráfico de vídeo en Internet, el uso de HTTP es su modo de descarga progresiva (*Progressive Download*) ha evolucionado a su uso mediante sistemas adaptativos (HAS, *HTTP Adaptive Streaming*). En el primer caso, los parámetros típicos para la evaluación de la QoE corresponden con el retardo inicial (dependiente del tamaño del buffer) y el número y duración de las interrupciones. Sin embargo, cuando se utiliza HAS, debido a que el cliente monitoriza y adapta el bit rate del contenido de los segmentos que se descarga, prácticamente no existen

interrupciones y adquiere más importancia el efecto en el cambio de calidad del contenido [4], [5].

Por otra parte, la simulación, evaluación e implementación de un testbed que representen escenarios más complejos que el transporte de Internet incrementa los retos relacionados con la evaluación de la QoE. Algunos de dichos escenarios constituyen las líneas de investigación del grupo de Comunicaciones Multimedia del iTEAM, como son las redes Ad Hoc, las redes híbridas para la transmisión a multitud de usuarios o el transporte de vídeo 3D.

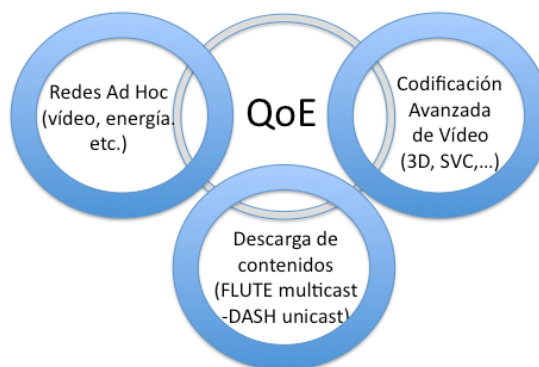


Fig. 1. Líneas de investigación sobre QoE.

II. QoE Y CONSUMO DE BATERÍAS EN REDES ADHOC

Las redes Ad Hoc, representan una alternativa en el desarrollo y soporte de nuevos escenarios de comunicación, propuestos principalmente como consecuencia de la gran difusión en el uso de dispositivos móviles. En particular, dicha popularidad ha dado lugar a planteamientos como por ejemplo el concepto de comunicación oportunista. Tal enfoque permitiría aprovechar la significativa densidad de dispositivos de usuario existente, por ejemplo en zonas urbanas, para el establecimiento de conexiones a través de un esquema colaborativo. O bien, ofrecer conectividad a Internet a través de la red Ad Hoc, como red de acceso [5].

Sin embargo tal potencial, aún debe hacer frente a desafíos propios de éste paradigma de comunicación. En particular, las limitaciones de energía de los dispositivos móviles debido al uso de baterías es un aspecto clave para su diseño. Si bien se han logrado avances en tecnología de baterías para incrementar su capacidad y tiempo de operación, el mayor inconveniente es que dichas mejoras no se han producido al mismo ritmo que la industria de dispositivos electrónicos.

Adicionalmente, el incremento de tráfico debido a la distribución de contenido multimedia y la mayor tasa de transmisión alcanzada con recientes estándares como IEEE 802.11n/ac contribuye a una operación intensa de las interfaces radio y un aumento drástico del gasto de energía [6]. La restricción de energía condiciona el tiempo de operación de un dispositivo en una red inalámbrica Ad Hoc y en consecuencia tiene un impacto negativo sobre la conectividad y la calidad de servicio en la transmisión de los flujos de datos.

En ese sentido, el medio inalámbrico es un aspecto clave de análisis en el gasto de energía. En particular, las zonas de cobertura comunes pueden generar un consumo notable e innecesario. Tal efecto tiene una relación directa con el principio de operación de las interfaces radio de los dispositivos, donde los estados de transmisión y recepción, son los modos de operación que mayor energía demandan.

Para conmutar a dichos estados, la interfaz de radio realiza un análisis del medio inalámbrico verificando los niveles de umbral de potencia en las señales existentes, ya sea para iniciar el modo de recepción o para detectar la disponibilidad del medio e iniciar una transmisión.

Tal mecanismo de operación ocasiona que en zonas comunes, los dispositivos conmuten a un estado de recepción debido a la detección de señales con un umbral de potencia adecuado, aún cuando no sean el objetivo de una transmisión. Dicho efecto denominado *overhearing* tiene un impacto significativo sobre el gasto de energía y a su vez se ve incrementado de forma crítica conforme la densidad de dispositivos se incrementa en el medio inalámbrico.

Una de las líneas de investigación del grupo COMM es el desarrollo de propuestas que permitan optimizar el gasto de energía en una red Ad-Hoc. En concreto se está trabajando en la implementación de modificaciones sobre uno de los protocolos más empleados en este tipo de redes como lo es el protocolo OLSR (*Optimized Link State Routing*). Las propuestas a nivel de red permiten alcanzar un compromiso entre los cambios requeridos en los nodos y las mejoras obtenidas por dichos mecanismos.

Hasta ahora, las modificaciones existentes emplean fundamentalmente el nivel de energía residual como métrica adicional para equilibrar el gasto de energía en la red. Sin embargo tales planteamientos tienen como principales inconvenientes, la sobrecarga de rutas y el enmascaramiento de nodos con niveles críticos de energía en la misma. Por tanto la diversificación de rutas basada únicamente en el nivel de energía residual en cada nodo intermedio, no es un mecanismo completo para lograr un equilibrio energético en la red.

Esto se debe a que se debería tener en cuenta el número de nodos vecinos, ya que dicho aspecto que depende de la ubicación de un dispositivo influye notablemente en el consumo de energía, debido al efecto *overhearing*. En particular los nodos con un gran número de dispositivos vecinos dentro de su área de cobertura tienden a ser sobre utilizados debido a la mayor probabilidad de alcanzar un destino a través del mismo. Dicho comportamiento genera un continuo gasto de energía para el nodo en cuestión y para todos sus vecinos dentro del área de cobertura común [7].

La propuesta se enfoca específicamente en alcanzar una planeación estratégica del gasto de energía, mediante una reducción de la carga de tráfico sobre aquellos nodos de mayor influencia en la red, en cuanto a los dispositivos

vecinos que tienen a su alcance. La menor operación de tales nodos en la conformación de rutas permitiría reducir significativamente el efecto *overhearing* sobre un mayor número de dispositivos. Adicionalmente la distribución de tráfico hacia rutas con menor densidad de nodos tiene el efecto positivo de una probabilidad menor de pérdidas de datos debido a la menor contienda por el medio.

Para alcanzar la planificación del gasto de energía descrito, se realiza una clasificación de los nodos presentes en la red asignándoles un valor estratégico acorde a su nivel de conectividad. Dicha métrica a continuación se computa en conjunto con el nivel de energía disponible en un nodo para el establecimiento de una ruta.

Para la implementación de la propuesta se ha trabajado a nivel de simulaciones en NS3, para lo cual se diseñó diversos escenarios de análisis. Los resultados obtenidos hasta el momento indican una mejora de la contribución (OLSR Estratégico, S_{OLSR}), frente a trabajos previos, mediante una reducción del gasto de energía y una extensión del tiempo de operación de la red, como se aprecia en la Figura 2.

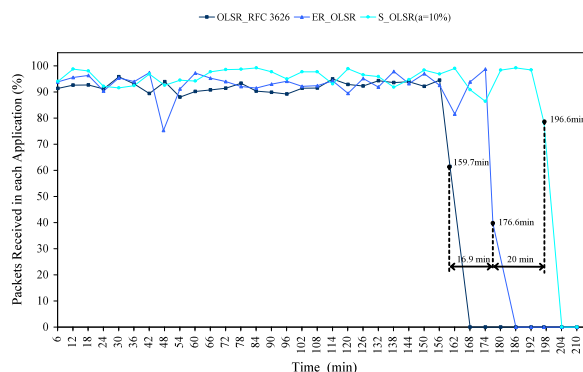


Fig. 2. Efecto de la distribución de consumo de energía sobre el tiempo de operación de la red.

En la Figura 3, se presenta un esquema de la metodología empleada para la evaluación.

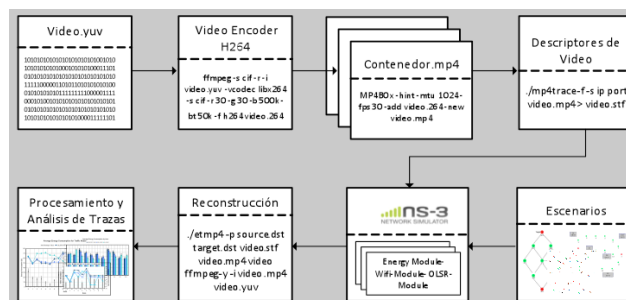


Fig. 3. Metodología para la Evaluación de la Propuesta.

Actualmente se está trabajando en la evaluación de la propuesta sobre equipamiento real, para lo cual se ha desarrollado un banco de pruebas conformado por 10 nodos Ad Hoc implementados mediante plataformas de desarrollo basadas en linux embebido (Raspberry Pi). En las siguientes figuras se puede ver el nodo implementado y su distribución en el campus de la UPV mediante una sesión de medidas. El siguiente paso en la investigación es validar las mejoras obtenidas en las simulaciones por el algoritmo propuesto

sobre un sistema formado por nodos reales, transmitiendo flujos de vídeo a lo largo de la red.



Fig. 4. Testbed real para la evaluación de la propuesta.

III. QoE Y DESCARGA DE CONTENIDOS

Otro escenario interesante de comunicaciones es la descarga de contenidos multimedia (vídeo, imágenes, ficheros, etc.) a través de redes inalámbricas en zonas con alta densidad de usuarios. Desde el punto de vista de los protocolos útiles para dicho escenario, destacan FLUTE (*File Delivery over Unidirectional Transport*) y DASH (*Dynamic Adaptive Streaming over HTTP*). En los últimos años, el estándar DASH [8] basado en el concepto HAS comentado anteriormente, junto con soluciones propietarias (Microsoft Silverlight Smooth Streaming, Apple HTTP Live Streaming y Adobe HTTP Dynamics Streaming), se han desarrollado con fuerza para la transmisión de vídeo. DASH es un estándar ISO (*International Organization for Standardization*) para la transmisión de contenido bajo demanda y en directo. DASH está basado en la segmentación de ficheros multimedia. Concretamente, en DASH cada fichero multimedia se codifica en diferentes calidades y cada fichero de calidad se divide en pequeñas porciones llamadas segmentos (con una duración de 1s a 15s). Para acceder al contenido multimedia, los clientes (mediante el método HTTP GET) seleccionan qué calidad de cada segmento quieren descargarse. DASH define un fichero de manifiesto, llamado MPD (*Media Presentation Description*), que describe el contenido multimedia, las diferentes calidades y cómo se realiza la división en segmentos. Cada servicio de vídeo corresponde con una presentación multimedia, que es una colección de ítems multimedia dependientes del tiempo, como se muestra en la parte inferior de la Figura 5. Las presentaciones multimedia se componen de una secuencia de periodos. Los periodos son intervalos de tiempo a lo largo de la duración

del vídeo, que no se pueden solapar. Cada periodo tiene diferentes alternativas de codificación, referidas como representaciones. Una de las características inherentes al usar DASH es su uso en escenarios basados únicamente en unicast. Sin embargo, existen situaciones donde el uso de multicast ofrece mejoras en cuanto eficiencia realmente importantes (eventos deportivos, conciertos, centros comerciales, etc.).

Uno de esos escenarios se presenta en la Figura 5. Existen muchos clientes en el área de servicio y el envío del vídeo a través de conexiones unicast puede causar congestión en el enlace. Para evitar esto, los servicios de vídeo utilizan tecnología de streaming híbrida multicast/unicast.

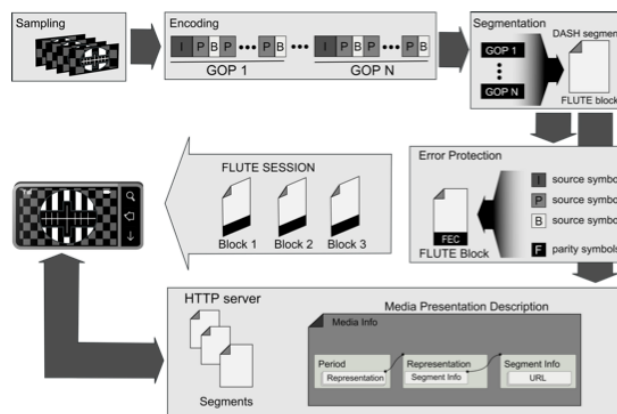


Fig. 5. Arquitectura híbrida FLUTE/DASH para el envío de vídeo.

Como protocolo para la difusión de contenidos en broadcast se utiliza FLUTE. FLUTE, definido en la RFC 6726 [9] es un protocolo para el envío unidireccional de ficheros sobre redes IP, especialmente adecuado para redes multicast. Entre sus principales características destaca la escalabilidad masiva y la fiabilidad en la transmisión. Para proporcionar seguridad en el envío, FLUTE utiliza distintos mecanismos, entre los que destaca el uso de codificación frente a errores FEC (*Forward Error Correction*). Entre los códigos estandarizados para FLUTE se encuentran Reed-Solomon, LDPC (*Low Density Parity Check*) y códigos Raptors. Con respecto al proceso de transmisión de contenidos en FLUTE, cada fichero se divide en bloques y a su vez cada bloque se divide en símbolos de codificación. Hay dos tipos de símbolos de codificación: símbolos fuente (obtenidos a través de los datos del propio fichero) y símbolos de paridad (generados a partir de un codificador FEC). En FLUTE los ficheros se envían a través de sesiones de envío (identificadas por una dirección IP y un número de sesión). A su vez, cada sesión de envío puede contener varios canales de transmisión FLUTE (identificados por un número de puerto). Para que los clientes sean capaces de identificar los ficheros que se están transmitiendo en un determinado canal así como sus características (por ejemplo, nombre, tamaño, codificación...) se utiliza la Tabla de Transferencia de Ficheros (FDT), que corresponde con un documento XML (*eXtensible Markup Language*) que se transmite en el canal de transmisión, y que tiene que ser descargado por los clientes antes de poder descargar ningún contenido.

Volviendo al sistema de transmisión de vídeo híbrido propuesto (mostrado en Fig. 5), después del proceso de codificación, el vídeo se divide de acuerdo a una política

determinada de segmentación, generando segmentos DASH y bloques FLUTE. La información referida a la segmentación DASH se indica en el MPD. Entonces, los clientes pueden acceder a los segmentos a través de una sesión de FLUTE y un servidor DASH. La flexibilidad en cuanto a la segmentación de un contenido tanto en FLUTE como en DASH permite combinar ambas tecnologías para descargar tramas de vídeo [10]. Cabe señalar que en FLUTE el proceso de codificación se lleva a cabo en cada bloque, por lo que diferentes bloques pueden utilizar diferentes códigos FEC. Así, es posible proporcionar protección diferente a cada bloque, por lo que se pueden utilizar técnicas UEP (*Unequal Error Protection*) de forma fácil. Esto es especialmente útil para la transmisión de vídeo, en los que existen tramas más importantes que otras.

Los clientes se conectan a la sesión FLUTE multicast y filtran los paquetes pertenecientes a los objetos que se quieren descargar. Así, los clientes reciben los diferentes símbolos que componen cada bloque y reconstruyen los GoPs correspondientes. Los símbolos de paridad recibidos se utilizan para reconstruir los símbolos fuente perdidos. Si después de realizar la decodificación FEC algunos paquetes no han podido ser reconstruidos, los clientes pueden solicitar los paquetes perdidos al servidor DASH a través de una conexión unicast. La manera en la que DASH realiza la segmentación es bastante flexible, así cada segmento puede contener desde una única trama de vídeo a varios GoPs.

Los clientes empiezan a reproducir el vídeo cuando reciben un cierto número de GoPs, para así tener una latencia inicial baja. Para proporcionar una reproducción continua del vídeo, el ancho de banda de transmisión utilizado tiene que ser igual o mayor que la tasa de reproducción de vídeo. Obviamente, si se dispone de un ancho de banda elevado, se podrá proporcionar mayor protección y los clientes recibirán el contenido con menos errores. De esta forma, existe un compromiso entre el ancho de banda utilizado y la calidad percibida por los clientes. Para evaluar dicho compromiso, se ha utilizado el vídeo “*Sintel*” codificado a 1080p y utilizando diferente protección FEC, niveles de pérdidas y ancho de banda. Los resultados obtenidos se pueden ver en la Fig. 6. Cada barra vertical representa el ancho de banda utilizado por una configuración específica FEC y contiene tres puntos que representan los valores de PSNR (*Peak Signal-to-Noise Ratio*) para cada porcentaje de pérdidas del canal (5%, 25% y 50%). Encima de cada barra, los valores entre paréntesis indican la tasa de codificación de las tramas I, P y B, en este orden. Mencionar que una menor tasa de codificación implica mayor paridad. En todos los casos, las tramas I tienen al menos el mismo nivel de protección que las tramas P y B, ya que las tramas I son más importantes.

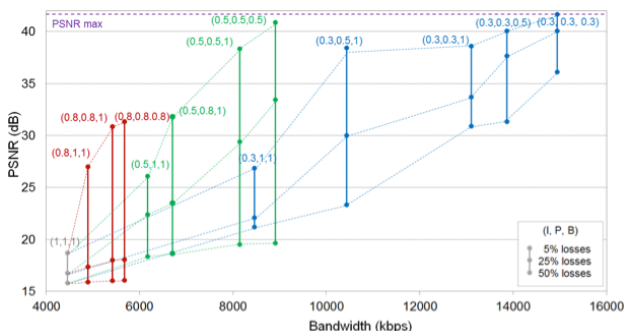


Fig. 6. Evaluación de la arquitectura híbrida FLUTE/DASH.

Como refleja la figura, incrementar el FEC de las tramas I mejora considerablemente el PSNR (*Peak Signal-to-Noise Ratio*), contrastando con lo que sucede con las tramas B, donde, en general, apenas mejora. Así, el uso de una alta protección no siempre representa una buena solución en términos de ancho de banda. Fijando un nivel de PSNR (directamente relacionado con el parámetro de calidad subjetiva MOS –*Mean Opinion Score*) y de ancho de banda, es posible elegir la configuración FEC óptima para un cierto porcentaje de pérdidas. Por ejemplo, en la Fig. 6, si un servidor quiere ofrecer un servicio de calidad excelente (un PSNR mayor de 35 dB) para clientes con un 5% medio de pérdidas y utilizando el mínimo ancho de banda, el servidor utilizará una configuración FEC de (0.5, 0.5, 1). Comparando con la configuración FEC de (0.3, 0.3, 1), esta última incrementa el PSNR en 0.27 dB a expensas de incrementar el ancho de banda un 61%, por lo que no resultaría una configuración óptima. En la Tabla 1 se muestra el incremento de la eficiencia al usar la arquitectura híbrida para diferentes situaciones de tasa de errores usando los valores de FEC óptimos y para diferentes valores de QoE medidos mediante el MOS. Utilizando valores de FEC adecuado, el uso de FLUTE para la descarga de contenidos y DASH para la recuperación de los segmentos perdidos, se obtiene un uso muy eficiente del ancho de banda y se garantiza una QoE (medido como PSNR) adecuada a los usuarios.

Tabla 1

MOS Level	Sintel		Tears of Steel	
	3 (Fair)	4 (Good)	3 (Fair)	4 (Good)
Scenario 1:				
34 users with 5% packet loss, 33 with 25% and 33 with 50%				
DASH recovery	1.9%	0.6%	8.5%	2.6%
Efficiency	92.0%	93.7%	85.4%	91.0%
Scenario 2:				
80 users with 5% packet loss, 10 with 25% and 10 with 50%				
DASH recovery	2.4%	0.7%	4.6%	2.8%
Efficiency	94.1%	96.2%	92.4%	93.6%
Scenario 3:				
10 users with 5% packet loss, 10 with 25% and 80 with 50%				
DASH recovery	2.0%	0.8%	13.2%	3.1%
Efficiency	89.0%	92.1%	72.1%	86.4%

IV. QOE Y CODIFICACIÓN DE VÍDEO 3D

Por último, y como tercer aspecto de investigación, se ha considerado un primer estudio de la codificación de vídeo 3D y la QoE. Las recientes mejoras en la tecnología de vídeo 3D han suscitado un creciente interés hacia el consumo contenidos de vídeo 3D. El número de pantallas de cine capaces de mostrar películas en 3D, así como el número de películas producidas en 3D ha aumentado constantemente en los últimos años. Con la disponibilidad de televisores con capacidad 3D y reproductores Blu-ray, y la introducción de los primeros canales de difusión en 3D se ha acercado el vídeo 3D a los usuarios, aumentando la popularidad y demanda del servicio, aunque comercialmente y como modelo de negocio a nivel usuario residencial todavía requiere un análisis de fondo.

Actualmente, la distribución de vídeo 3D se soporta en las redes DVB (*Digital Video Broadcasting*), y su reproducción está sujeta a la capacidad de los terminales para decodificar el vídeo 3D. Los sistemas de distribución de

vídeo 3D utilizan los formatos *frame-compatible* o *full-resolution frame-compatible* [11] para ser compatibles con los sistemas de distribución 2D, evitando con esto tener que realizar actualizaciones masivas de infraestructuras de transmisión por parte del proveedor y hardware de recepción por parte del usuario. En *frame-compatible* las dos imágenes (izquierda y derecha) resultantes del proceso de filtrado y muestreo se multiplexan en una trama MPEG-2 Transport Stream (TS) junto con la información de señalización y protección de errores correspondiente. El principal inconveniente de este formato es la reducción de la calidad de la imagen debido al uso de técnicas de diezmando y de interpolación. Por su parte, *Full-resolution frame-compatible* mantiene la calidad de ambas vistas, pero requiere el doble de ancho de banda. Junto a los formatos de vídeo estereoscópico ya mencionados, han aparecido los formatos multivista basados en las nuevas técnicas de codificación Multiview Video Coding (MVC) [12] y Multiview Video Plus Depth (MVD) [13], que permiten el envío de múltiples vistas requeridas entre otros casos, en los sistemas basados en pantallas auto-estereoscópicas. Los formatos multivista aumentan el ángulo de visión en 3D de los usuarios y también la eficiencia en el proceso de codificación en función del ancho de banda requerido.

El siguiente paso en la evolución de la distribución de vídeo 3D es su transporte de forma más eficiente a través de Internet. Como plataforma de distribución Internet ofrece muchas ventajas en comparación con la baja flexibilidad de las redes DVB. Por ejemplo, Internet puede permitir el envío de múltiples vistas, teniendo en cuenta las características heterogéneas de los dispositivos de reproducción y el ancho de banda disponible para cada usuario final. Sin embargo, muchos aspectos se deben tener en cuenta aún para que Internet se convierta en la plataforma por defecto para la distribución de vídeo 3D. En este punto el diseño cross-layer surge como tema principal de la investigación. Considera el proceso de generación y codificación de contenidos 3D, la monitorización del estado de la comunicación, la ejecución de algoritmos adaptativos, y la evaluación de la calidad de experiencia (QoE).

En el contexto de un sistema de distribución adaptativo de vídeo 3D a través de redes IP, el objetivo de la línea de investigación ha sido realizar una comparación de la calidad de vídeo 3D/estereoscópico obtenida empleando el estándar de codificación de vídeo 3D MVC H.264-MVC, el ya popular Advanced Video Coding H.264-AVC y el recientemente publicado High-Efficiency Video Coding H.265-HEVC.

Al evaluar la eficiencia de un codificador se debe considerar que es importante comprimir la señal de vídeo tanto como sea posible y también mantener la calidad de vídeo cerca de la original. Por lo general, encontramos dos categorías o métodos de medición de la calidad de vídeo: 1) la calidad subjetiva que cuantifica la calidad del vídeo usando evaluadores humanos; y 2) la calidad objetiva que se mide utilizando modelos matemáticos que pueden ser automatizados. La evaluación subjetiva de la calidad de vídeo resulta costosa en términos del tiempo requerido y la necesidad de evaluadores humanos, además, puede resultar ineficiente en escenarios que involucren aplicaciones en tiempo real. Por esta razón, la métricas de calidad objetiva pueden ser una alternativa adecuada en un escenario de transporte adaptativo, donde es deseable tener una referencia

de la calidad percibida por los usuarios a efectos de optimizar la red.

Mucho trabajo se ha hecho en la evaluación de la QoE para la reproducción de vídeo 2D y también unos cuantos trabajos para reproducción de vídeo 3D. De forma general seguimos los aspectos típicos para evaluar la QoE: evaluación objetiva [14] y evaluación subjetiva [15].

La evaluación objetiva de la calidad de vídeo 3D se basa en parámetros conocidos ya utilizados para el vídeo 2D como: Peak Signal-to-Noise Ratio (PSNR) y Structural Similarity (SSIM) Index. La implementación de las métricas de evaluación objetiva es simple, tiene bajo coste computacional y las medidas pueden reproducirse. Tanto el PSNR como el SSIM son métricas del tipo Full-reference, es decir se basan en la disponibilidad de la señal original, que es contrastada con la señal degradada, frame a frame.

El índice SSIM es un método diseñado para medir la similitud entre dos imágenes. SSIM mide la calidad de la imagen decodificada mediante la comparación frame a frame con una imagen sin comprimir o sin distorsión inicial como referencia. SSIM está diseñado para mejorar los métodos tradicionales, como el PSNR, que han demostrado ser incompatibles en determinadas situaciones con la percepción del ojo humano. SSIM considera que variaciones en la luminosidad y contraste pueden no afectar mucho a la calidad de la imagen, obteniendo valores más cercanos a la QoE del usuario.

La evaluación subjetiva de vídeo 2D puede ser utilizada también para reproducción de vídeo 3D. Este tipo de evaluación está centrada en el usuario, lo cual es importante dado que puede destacar aspectos relativos a sus requerimientos. Los métodos más usuales full-reference o no-reference estandarizados en ITU-R BT.500 [16] son: Double-Stimulus Impairment Scale (DSIS), Double-Stimulus Continuous Quality-Scale (DSCQS), Simultaneous Double Stimulus for Continuous Evaluation (SDSCE), Absolute Category Rating (ACR), Hidden References (HRR), ACR5, ACR5-HRR, ACR11 y ACR11-HRR. Un test basado en el método full-reference compara el vídeo codificado frente al original no comprimido y comprueba las diferencias. Una técnica no-reference analiza el vídeo bajo prueba y busca artefactos sin realizar ninguna comparación. Ambas aproximaciones al problema presentan puntos fuertes y débiles, y escoger la aproximación correcta depende de la aplicación pretendida, las limitaciones de coste, y otros aspectos prácticos.

El sistema visual humano es extremadamente complejo, y puede detectar fácilmente algún tipo de distorsión, mientras que puede pasar por alto otras, dependiendo de diversos factores. Estos factores pueden incluir el tipo de aplicación (TV, vídeo conferencia, etc.), el lugar de la imagen en donde se produce la degradación (generalmente las degradaciones son menos visibles en regiones con muchos detalles o "actividad espacial", o con gran movimiento, y son más visibles en imágenes estacionarias, o en fondos poco texturados). Incluso la calidad percibida puede depender del tipo de dispositivo utilizado y del tamaño del monitor. Las desventajas de la evaluación subjetiva son sus requisitos de: ambiente controlado, equipamiento específico y un número suficiente de "beta testers" (usuarios finales dispuestos a realizar los tests).

La Figura 7 muestra el marco de trabajo para el estudio comparativo de la calidad de vídeo 3D usando los estándares

de compresión H.264-AVC, H.264-MVC y H.265- HEVC. También se ha considerado la ventaja del video 3D asimétrico. Dicha ventaja reside en poder reducir la calidad de una de las vistas y mantener la calidad en la otra vista. La QoE del usuario corresponderá con la QoE asociada a la vista de mejor calidad.

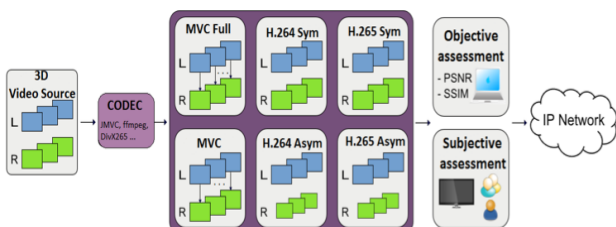


Fig. 7. Esquema general estudio comparativo de la calidad de video 3D.

Como señal de prueba se ha utilizado el video de ejemplo “Oldtimers” de Werner Bloos 1080p Dual files (Figura 8). Se dispone de la vista derecha e izquierda con una resolución espacial de 1440 x 1080 pixels con una tasa de 25 fps y fue grabado utilizando una configuración dual sincronizada Sony HDR-FX1.

Se trata de una película corta grabada al aire libre con iluminación natural, presenta un movimiento moderado de la cámara, con cambios de escena que ofrecen diferentes características de iluminación y varios tipos de objetos en movimiento a diferentes velocidades. Por esta razón consideramos que los resultados del análisis llevado a cabo pueden extrapolarse a casi cualquier secuencia con mejores o similares características.



Fig. 8. Secuencia de test de video 3D.

Como entrada a los codificadores las secuencias son provistas en formato YUV 4:2:0 con una duración de 25 s lo que es equivalente a 625 frames. Así mismo para utilizar el tamaño de video apropiado a la resolución de pantalla de los dispositivos móviles, la resolución especial original fue reducida a 400x240 pixeles.

Como criterio para la selección de las herramientas de codificación a emplear se tuvo en cuenta la estabilidad, la flexibilidad en términos de la posibilidad de modificar los parámetros de configuración, y su carácter abierto o propietario, entre otros. Con el objetivo de obtener resultados comparables entre los diferentes estándares, se eligieron parámetros de configuración similares entre los codificadores. Para medir la eficiencia de codificación,

cuatro diferentes parámetros de cuantificación (QP) han sido utilizados 24, 30, 36 y 42.

Como se ha comentado, en este trabajo se han seleccionado PSNR y SSIM como métricas objetivas de calidad. Se utilizó *Evalvid Tools* para calcular estos parámetros para cada frame de la secuencia de video. Los resultados de este proceso se presentan en la Figura 9 y Figura 10. En Figura 9 se muestra la comparativa del PSNR en función de la tasa de bits entre H.264-MVC, H.264-MVCFull, H.264-AVC y H.265-HEVC es sus versiones simétrica y asimétrica. Mientras, la Figura 10 muestra la comparativa del SSIM para las mismas secuencias de vídeo. La tasa de bits fue medida en kbps, la PSNR en dB y el SSIM dentro del intervalo [0, 1].

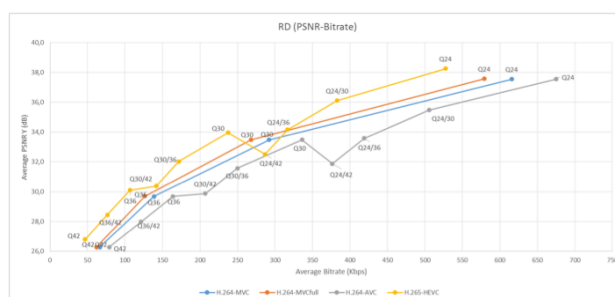


Fig. 9. PSNR promedio vs Bitrate

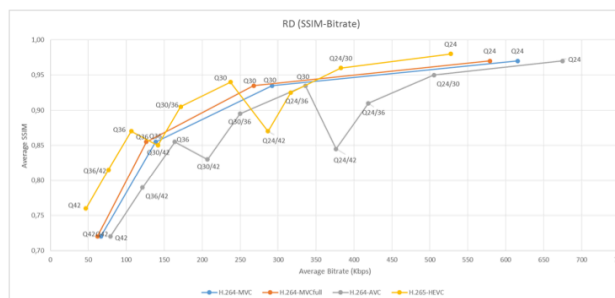


Fig. 10. SSIM promedio vs Tasa de bits

De la curva en Figura 10 se observa que: H.265-HEVC es más eficiente como esquema de codificación en función del PSNR respecto a H.264 en cualquiera de sus implementaciones (H.264-AVC, H.264-MVC y H.264-MVCFull); el aumento de QP tiene un efecto directo sobre la reducción de la calidad de la secuencia de vídeo tanto en función del PSNR como del SSIM; y la codificación H.264-MVCFull gracias a su mayor nivel de predicción entre vistas, resulta más eficiente comparada respecto a H.264-MVC y H.264-AVC que fueron codificadas con el mismo codificador JMVC 8.5 y los mismos parámetros de configuración.

Aunque las métricas de evaluación objetiva PSNR y SSIM, son ampliamente utilizadas, solo proporcionan una aproximación a la calidad percibida por el usuario. En consecuencia el paso siguiente consiste en la realización de test subjetivos que permitan validar la calidad objetiva en relación a la calidad de la experiencia.

En este trabajo la evaluación subjetiva se ha llevado a cabo con base en la recomendación BT.500 mediante una variación del método de medida de calidad de doble estímulo (DS) con una referencia explícita y una escala Mean Opinion Score (MOS) de 1 a 5, donde 5 corresponde con una calidad excelente y uno a una mala calidad.

Se evaluaron subjetivamente 28 secuencias de video de 25 segundos a 25 fps, generadas usando cuatro métodos de codificación diferentes y cuatro parámetros de cuantificación en cada caso: H.265-HEVC Simétrico/Asimétrico (10), H.264-AVC Simétrico/Asimétrico (10), H.264-MVC (4) y H.264-MVCFull (4). Para facilitar la tarea de evaluación y evitar la fatiga durante la aplicación del test, las secuencias a evaluar se distribuyeron en cuatro sesiones. Durante cada sesión a los evaluadores se les pidió valorar 7 secuencias de video reproducidas en orden aleatorio. Antes de cada conjunto de secuencias estaba precedido por un video de referencia, correspondiente a la señal original sin ninguna degradación, y un video de mala calidad. De esta forma los participantes conocían el rango de variaciones de calidad que se podían encontrar durante el test. Cada secuencia de video fue evaluada individualmente justo tras ser presentada al usuario, que tuvo entre 5 y 10 segundos para valorar su QoE de forma general en términos de la sensación de profundidad y la calidad espacio-temporal percibida. Los videos fueron reproducidos a través del reproductor VLC que desde noviembre de 2013 en su versión 2.1.0 introdujo soporte para la decodificación de H.265-HEVC. Para su visualización se emplearon unas gafas de visión 3D anáglifo y un monitor LCD de 17 pulgadas. El formato de video empleado para la presentación de los datos al reproductor fue el frame packing side-by-side.

Otro aspecto importante a tener en cuenta es la selección de los evaluadores para la realización del test. Un total de 20 evaluadores en un rango de edad entre los 21 y los 40 años, con una edad promedio de 28 años, con buena agudeza visual y estereoscópica fueron seleccionados. Para el test subjetivo de video, se tomaron dos grupos de evaluadores. Un grupo de evaluadores no expertos, en el sentido que ninguno estaba directamente relacionado en análisis o evaluación de video 3D en su trabajo cotidiano y no tenían experiencia en la evaluación de video ni imagen y un segundo grupo de evaluadores con experiencia en la temática de QoE. Las puntuaciones proporcionadas por los evaluadores fueron promediadas para obtener la puntuación MOS. La Figura 11 muestran los resultados obtenidos.

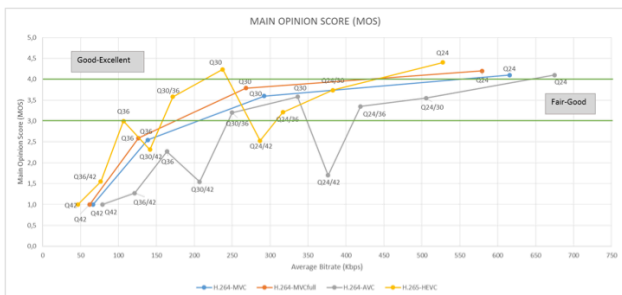


Fig. 11 Puntuaciones de tests subjetivos para MVC, MVC-Full, H.264-AVC Simétrico/Asimétrico y H.265-HEVC Simétrico/Asimétrico.

Con base en la Fig.11 observamos que sobre la referencia de un nivel de calidad equivalente en el rango *Good-Excellent* el estándar de codificación H.265-HEVC presenta

un ahorro de un 64.81% en función de la tasa de bits respecto a la codificación con H.264-AVC, un 61,44 % respecto a H.264-MVC y un 59% respecto a H.264-MVCfull. También se observa el porcentaje de ahorro en función de la tasa de bits para un nivel de calidad equivalente en el rango de calidad *Fair-Good* en el que la secuencia H.265-HEVC VLeft QP30/VRright QP36 es la que reporta el mejor compromiso de calidad respecto a la tasa de bits.

También se comprueba que una alta degradación de la vista derecha codificada con QP=42 hace que la calidad percibida por el usuario se considere inaceptable incluso al combinarla con una vista izquierda de muy buena calidad.

De acuerdo a la Figura 11, para los codificadores H.264-AVC Simétrico/Asimétrico y H.265-HEVC Simétrico/Asimétrico, se observa que se dispone de 5 niveles para la adaptación de la tasa de bits en los rangos de (674,88 a 249,76 Kbps) y (527,68 a 171,68 Kbps) respectivamente, manteniendo un nivel de calidad por encima de aceptable (MOS>3). Por su parte, en el caso de la codificación basada en H.264-MVC y H.264-MVCFull se dispone únicamente de dos niveles de adaptación en los rangos (615,84 a 292,16 Kbps) y (579,2 a 697,84 Kbps).

Por su parte los resultados de la calidad subjetiva para las codificaciones H.264-MVC respecto a simulcast H.264-AVC tomando como referencia una misma tasa de bits, muestran que las implementaciones basadas en H.264-MVC, reportan mejor calidad visual que las simulcast, siendo esto coherente con los resultados encontrados en la bibliografía.

El siguiente paso en el estudio, es analizar la QoE considerando pérdidas y retardo, asociado a las redes móviles o Internet.

V. CONCLUSIONES

En este artículo se han descrito brevemente algunas de las líneas de investigación del grupo COMM del iTEAM en relación al concepto de QoE. Dicha investigación cubre desde la evaluación de nuevos parámetros, como los cambios de calidad en streaming basado en HAS, al estudio del consumo de energía en una red Ad Hoc para ampliar el tiempo de vida de la propia red; la evaluación de la arquitectura híbrida basada en FLUTE y DASH; o la evaluación de la QoE respecto a la codificación de video 3D.

VI. AGRADECIMIENTOS

Este artículo ha sido realizado en el grupo de Comunicaciones Multimedia del iTEAM-UPV, y con el apoyo de la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación- Gobierno de Ecuador (Beca 195-2012) y el proyecto PROMETEOII/2014/003 financiado por la Generalitat Valenciana.

REFERENCIAS

- [1] Y. Chen, K. Wu, Q. Zhang, "From QoS to QoE: A Tutorial on Video Quality Assessment", aceptado en IEEE Communications Surveys & Tutorials, DOI 10.1109/COMST.2014.2363139, 2014.
- [2] P. Juluri, V. Tamarapalli, D. Medhi, "Measurement of Quality of Experience of Video-on-Demand Services: A Survey", aceptado en IEEE Communications Surveys & Tutorials, DOI 10.1109/COMST.2015.2401424, 2015
- [3] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2013-2018" Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

- [4] M. Seufert et al., "Survey on Quality of Experience of HTTP Adaptive Streaming", IEEE Communications Surveys & Tutorials, Vol. 17, No.1, First Quarter, 2015.
- [5] Marco Conti, Silvia Giordano. (2014, "Mobile Ad Hoc Networking: Milestones, Challenges, and New Research Directions", Communications Magazine, IEEE, Vol. 52, Issue 1, pp. 85-96, 2014.
- [6] S. González, P. Arce, and J. C. Guerri, "Técnica de buffering para la optimización del consumo de energía en la transmisión de tráfico multimedia en redes ad-hoc," IEEE Latin America Transactions, vol. 13, no. 1, pp. 250-258, 2015.
- [7] S. González, P. Arce, and J. C. Guerri "Nodos estratégicos: una propuesta para mejorar la eficiencia energética en redes inalámbricas Ad-Hoc," presented at Simposium Nacional de la Unión Científica Internacional de Radio (URSI), Valencia (Spain), Sep. 2014.
- [8] ISO/IEC 23009-1, "Dynamic adaptive streaming over HTTP (DASH) – Part 1: media presentation description and segment formats," 2012.
- [9] T. Paila, R. Walsh, M. Luby, V. Roca and R. Lehtonen, "FLUTE – File delivery over unidirectional transport," IETF RFC 6726, 2012
- [10] R. Belda, I. de Fez, F. Fraile, P. Arce, and J. C. Guerri, "Hybrid FLUTE/DASH video delivery over mobile wireless networks," Transactions on Emerging Telecommunications Technologies, vol. 25, no. 11, pp. 1070-1082, 2014.
- [11] C. G. Gurler, B. Gorkemli, G. Saygili, A. M. Tekalp, Flexible Transport of 3-D Video Over Networks, Proceedings of the IEEE, Apr. 2011, vol. 99, no. 4, pp. 696-707.
- [12] A. Vetro, T. Wiegand, y G. J. Sullivan, Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard, Proceedings of the IEEE, Apr. 2011, vol. 99, no. 4, pp 626-642.
- [13] P. Merkle, A. Smolic, K. Müller y T. Wiegand, Multiview Video Plus Depth representation and coding Image Processing, 2007. ICIP 2007. IEEE International Conference, 2007, Vol. 1, pp 201-204.
- [14] S. Chikkerur, V. Sundaram, M. Reisslein, L. J. Karam, Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison, IEEE Transactions on Broadcasting, Jun. 2011, vol. 57, no. 2, pp. 165-182.
- [15] P. Aflaki, M. M. Hannuksela, J. Hakkinen, P. Lindroos, and M. Gabbouj, Subjective study on compressed asymmetric stereoscopic video, in Proc. IEEE Int. Conf. Image Process., Hong Kong, Sep. 2010, pp. 4021–4024.
- [16] ITU-R Recommendation BT.500-13, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, Geneva, 2012.

Estimación Proactiva de QoS de Canales Inalámbricos para Mejorar la QoE en Servicios Multimedia

Álvaro Suárez^{2,3}, Elsa Macías López^{2,3}, Vicente Mena^{1,3}, Domingo Marrero^{2,3}, José A. Santana^{1,3}

¹Departamento de Señales y Comunicaciones
{jose.santana,vicenteeingenio.mena}@ulpgc.es

²Departamento de Ingeniería Telemática
{alvaro.suarez, elsa.macias, [domingo.marrero](mailto:domingo.marrero@ulpgc.es)}@ulpgc.es

³Grupo de Arquitectura y Concurrencia (GAC)
Universidad de Las Palmas de Gran Canaria
Campus Universitario de Tafira s/n
35017 Las Palmas de Gran Canaria

Resumen- Los servicios multimedia IP llegan a una gran mayoría de hogares, hoteles y otros establecimientos de ocio (grandes cadenas comerciales, pabellones de deporte y otros espectáculos...) que son visitados regularmente por una masa de personas comunicadas continuamente con sus terminales móviles dotados de capacidad de comunicación multimedia. Aunque las troncales de comunicación del Núcleo de Internet soporta velocidades astronómicas y los enlaces de bajada sobre fibra óptica de última milla son muy elevadas (300 Mbps), los canales de subida (30 Mbps o 200 Mbps como máximo y sólo en ciertos casos), limitan el despliegue de los servicios multimedia severamente a la gran masa. Pero aún más, limitan este despliegue, los enlaces inalámbricos de última milla, que aunque pueden ser de alta velocidad, tienen problemas inherentes que degradan severamente la calidad de servicio y la calidad de experiencia de usuario. En este trabajo hacemos una revisión breve del trabajo que ha realizado el Grupo de Arquitectura y Concurrencia en los últimos años para paliar esta problemática y terminamos exponiendo brevemente los últimos avances que hemos logrado para proveer a los servicios multimedia de telecomunicación: telefonía y televisión, con mecanismos robustos que les permitan mejorar la calidad de experiencia de usuario.

Palabras Clave- RSSI, Wireless Fidelity, Estimación, Medias Móviles, Aplicaciones Inalámbricas, Multimedia, QoE, QoS.

I. INTRODUCCIÓN

Los servicios multimedia como tales tienen su origen en los antiguos servicios telefónicos y de Televisión. Desde que Antonio Meucci, inventó el *teletéfono* (con tan mala suerte que, a falta de recursos para patentar el invento, lo entregó a una Empresa que ni siquiera le devolvió los materiales) [1], hasta el uso del proyecto *Web Real time Communications (WebRTC)* para las empresas [2], se han utilizado muchas formas de llevar el servicio al usuario final. La tendencia es a desplegar la telefonía sobre *Internet Protocol (IP)*.

Por lo menos Paul Gottlieb Nipkow si pudo patentar el disco de Nipkow que implementaba conceptos de fototelegrafía, origen de los sistemas mecánicos de televisión

de los años 1930 [3]. Desde entonces, se han utilizado muchas tecnologías, redes y servicios para hacer llegar la televisión al usuario final. Un elemento importante es que el servicio original de televisión ha sufrido muchas mutaciones que la telefonía, inventándose a lo largo de los años muchos tipos de servicios de emisión de vídeo, como el vídeo bajo demanda (*VoD*, del inglés *Video on Demand*), el vídeo en vivo, sistemas de podcasting, vodcasting y broadcatching, la televisión multi-pantalla, la televisión híbrida... La tendencia igualmente es emitir todos estos servicios usando IP.

Los servicios de telefonía [2] y de televisión [4] basados en La nube pueden revolucionar definitivamente estos servicios en un futuro cercano, en los que con terminales móviles muy simples se podría acceder a dichos servicios. Esto supondría un avance técnico (el despliegue en la Nube del servicio virtualizado de televisión y telefonía), económico (millones de usuarios con bajos recursos económicos podrían usar el servicio adquiriendo terminales baratos) y de modelo de negocio (y las barreras para establecer operadores virtuales de telefonía y televisión podrían derribarse definitivamente).

En el marco de Internet, varios eventos importantes han ocurrido a finales del siglo XX hasta ahora: las redes sociales han propiciado el desarrollo de nuevos modelos de televisión (televisión social), y la descarga masiva de videos bajo demanda que descargan los amigos. Por otro lado, en las redes sociales, hoy día se proveen servicios de mensajería que pretenden ser una forma alternativa a la telefonía tradicional. Aplicaciones de vídeo en servicios de *microbloging* de uso masivo: al igual que los mensajes de no más de 140 caracteres, ahora está de moda el envío de videos de muy corta duración mediante herramientas como *periscope*. Las redes P2P también han provocado una explosión de nuevos servicios de telefonía y de vídeo de películas que se descargan sin tener que pasar por un servidor central.

Todos estos servicios tienen que llegar hasta el usuario final. Los terminales preferidos por los usuarios actualmente

para acceder a estos servicios son sus terminales móviles que pueden usar *Long Term Evolution (4G)* o bien *Wireless Fidelity (WiFi)*, para hacer la comunicación. El uso de estas tecnologías impone serías restricciones a la provisión de estos servicios. Aunque la velocidad de acceso cada año mejora, lo que no mejora tanto es la Calidad de Servicio (*QoS*, del inglés *Quality of Service*) de la Red. En concreto, para los servicios multimedia, estas redes tienen el problema que provocan interrupciones repentina e intermitente de los terminales inalámbricos. Nosotros hemos estudiado ampliamente este fenómeno y presentado algunos métodos para mitigar su efecto en servicios multimedia [5].

En el apartado II presentamos en general las causas que producen este problema. En el apartado III presentamos algunas ideas que hemos trabajado en estos momentos. Finalmente presentamos unas breves conclusiones y posibles ideas de trabajo futuro.

II. LAS INTERRUPCIONES DE SERVICIO MULTIMEDIA EN REDES WIFI

Las redes WiFi se usan hoy en día para conectar personas que encuentran un único *Punto de Acceso (PA)* y deben usarlo para acceder a Internet. En este caso es habitual que los usuarios puedan experimentar interrupciones del servicio, cuando, entre otros:

- El terminal experimenta una variación del *Received Signal Strength Indicator (RSSI)* no esperado. Esto es debido a obstáculos, o bien que el terminal en movimiento sale y entra en zonas de cobertura variable.
- Existen muchos usuarios conectados y comunicando tráfico muy pesado en el mismo PA.
- Existen otros PA que interfieren el canal.
- El *Algoritmo de control de Acceso al Medio (MAC)* no funciona como debiera.

La planificación meticulosa y cuidada de las redes inalámbricas no es la solución a este problema: siempre puede ocurrir que un número de usuarios no controlados a priori puedan estar interfiriendo en el canal en un momento dado. Con el aumento de la velocidad en estas redes, es cierto que este problema se puede mitigar; pero no solucionar definitivamente. Además, en el típico edificio de viviendas en el que no existe control sobre los PA de los vecinos, y además los PA son de muy bajo coste y el conocimiento de los vecinos no es elevado, dimensionar adecuadamente la red no suele ser posible.

Otro caso en el que se pueden producir interrupciones de servicio es en aquellas redes en las que existe más de un PA al que conectarse. Esta situación se suele dar en los campus universitarios, centros comerciales, instalaciones deportivas, hoteles... Normalmente los PA suelen interconectarse entre ellos mediante Ethernet y un servidor típico que proporciona control de admisión a Internet. En este caso, el problema típico es el *handover* horizontal en WiFi. Este problema ha sido ampliamente analizado para tráfico de *Voz sobre IP (VoIP)* y *VoD*. Sin embargo, todavía sigue existiendo problemas que no se han resuelto definitivamente. Nosotros hemos estudiado el problema del *handover* vertical entre WiFi y WiMAX en [5]. Se pueden resolver cuando la

velocidad del terminal es muy baja (usuario caminando a baja velocidad normalmente) y la carga de la red es baja o bien se provee algunos nodos estáticos (*malla*) que ayudan a mitigar el problema. Actualmente, con el proyecto Fi de Google este problema ha vuelto a tomar importancia [6].

Otro escenario más complicado es el de las redes WiFi *ad hoc* planas o jerárquicas en la que todos los nodos se mueven constantemente. En este caso se debe proveer un algoritmo de encaminamiento capaz de proactivamente descubrir problemas en los caminos para evitar que existan interrupciones de servicio [7]. A pesar de que las redes *ad hoc* no han tenido el impacto práctico que se esperaba, desafíos interesantes a día de hoy es aplicar estas técnicas a redes vehiculares en las que los terminales se pueden mover a gran velocidad y en redes de sensores multimedia.

Otro escenario en el que se pueden producir las interrupciones de servicio son en las redes 4G que utilizan técnicas de *offloading* usando WiFi como red auxiliar para cursar el tráfico a Internet.

Por último en las redes 5G WiFi es una de las tecnologías que se puede usar como red de acceso. Según [8], para manejar la complejidad de la interoperación de tecnologías de acceso, se debe incluir procedimientos de red inteligentes, que manejen adecuadamente niveles de interferencia, movilidad, información sobre el tráfico, entre otros. Nosotros añadimos que además se debe proveer métodos adecuados para mitigar las interrupciones de servicio multimedia.

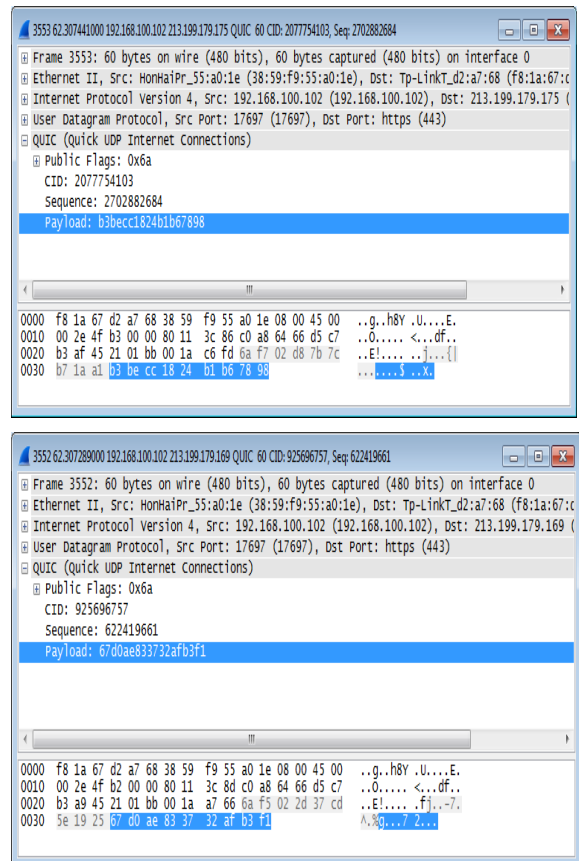


Fig. 1. Utilización de QUIC en Skype.

Podríamos pensar que las interrupciones de servicio son únicamente debidas a WiFi; pero recordemos que la provisión de los servicios multimedia se está haciendo sobre IP. En ese caso, hay otras fuentes de problemas que también causan posibles interrupciones. Entre otras, podemos mencionar que el rendimiento del *User Datagram Protocol (UDP)* no es adecuado porque no permite hacer un control efectivo a través de *sockets* sobre la aplicación del terminal inalámbrico. Además, el uso del *Transmission Control Protocol (TCP)* da problemas sobre redes inalámbricas debido a que no controla bien la pérdida de paquetes con los errores de canal producidos inherentes a este tipo de redes. A día de hoy, a nivel comercial, se suelen utilizar otros protocolos propietarios como *Quick UDP Internet Connections (QUIC)*, que se muestran en la Fig. 1, que tratan de resolver este tipo de problemas.

Destacar que el uso de *Hyper Text Transfer Protocol (HTTP)*, sobre el que se puede hacer directamente la comunicación de video y audio actualmente, no es capaz de proporcionar mecanismos de control eficaces sobre redes inalámbricas para evitar las interrupciones. Con la llegada recientemente de HTTP 2.0 es de esperar que este escenario pueda cambiar.

III. ALGUNAS IDEAS RECIENTES

En este apartado presentamos los experimentos que hemos realizado con VoIP y video streaming en redes inalámbricas, y por último presentamos las ideas que estamos desarrollando para mitigar estos problemas con nuevas propuestas.

A. Experimentos con VoIP en redes inalámbricas

La idea es presentar medidas reales del tráfico de VoIP en un par de terminales inalámbricos para demostrar los problemas reales de este servicio en estas redes.

El escenario fue constituido por 2 terminales, que establecieron una comunicación en modo *Ad hoc* utilizando el estándar IEEE 802.11g. Se escogió un lugar donde no hay prácticamente ningún PA y se hicieron las pruebas desde las 10:00 hasta las 13:00, con clima soleado, y una temperatura promedio de 29°C. Como los sistemas operativos de los teléfonos móviles, hoy día no permiten implantar redes *ad hoc* fácilmente, se utilizaron dos computadores portátiles. En uno de ellos se virtualizó un servidor de VoIP Asterisk, con la tarjeta de red de la máquina virtual configurada en modo bridge. En este equipo también se instaló un softphone Zoiper que tenía asignado una de las 2 extensiones configuradas sobre el servidor, ambas extensiones eran sobre *Session Initiation Protocol (SIP)*. La otra extensión fue configurada en un softphone Zoiper de otro computador portátil de la red *ad hoc*.

Los computadores personales fueron colocados inicialmente a una distancia de 1 m, se encontraban sobre una superficie plana a una altura de 45 cm del nivel del suelo. La distancia de separación fue aumentando en un intervalo de un metro hasta llegar a 18 m, luego fue aumentando 2 metros hasta tener una distancia de 30 m. Para aplicaciones *Ad hoc* esta distancia es suficiente. Después de cada movimiento de

terminal se ejecutaba la desconexión de la red y la conexión en cada distancia de medición.

Las características técnicas de las tarjetas de red se muestran en las tablas I y II.

Utilizamos software para medir el RSSI (Vistumbler, Commview e Inssider) y para medir el tráfico (Iperf y Wireshark).

Entre los dos computadores siempre iniciábamos una conversación de VoIP. Siempre era la misma conversación, para tener controlado el mensaje y la cantidad de paquetes que se enviaban. La idea era poder obtener conclusiones sobre la influencia de distintos parámetros de QoS y observar si el usuario experimentaba interrupciones importantes en la conversación.

Con el software anterior hicimos medidas de: jitter, paquetes perdidos, errores de secuencia *Real Time Protocol (RTP)*, tasa de transferencia media, y RSSI, todos en un solo sentido. A partir de ahí se observó la correlación existente entre los niveles de RSSI y las pérdidas de paquetes, entre la *Relación Señal a Ruido (SNR)* y la distancia entre los computadores, el RSSI y la distancia, el RSSI y el SNR, el jitter y el SNR, pérdida de paquetes y SNR (RSSI), pérdida de paquetes y la pérdida de secuencia de RTP. También se estudió el efecto del CODEC G.711.

En la Tabla III se muestran los coeficientes de correlación de Pearson obtenidos al analizar estadísticamente las variables medidas. La idea es mostrar la correlación que existe entre las variables estudiadas, para ver si existe alguna relación entre ellas. Se pudo observar las típicas relaciones del RSSI y el SNR con la distancia [m]. Además se observó la correlación lineal entre el SNR y el RSSI. Por otro lado se pudo observar la relación entre el jitter y las pérdidas de paquetes.

Tabla I
CARACTERÍSTICAS DEL PRIMER COMPUTADOR PERSONAL

Procesador	Intel® Core™ i5-3210M CPU 2.50GHz
RAM	4 GB
WiFi	IEEE 802.11n, IEEE 802.11g, IEEE 802.11b, 802.11d, 802.11g

Tabla II
CARACTERÍSTICAS DEL SEGUNDO COMPUTADOR PERSONAL

Procesador	Intel® Core™ i5-3230 CPU 2.60GHz
RAM	6 GB
WiFi	802.11a, 802.11b, 802.11g, 802.11n, 802.11d, 802.11e, 802.11i, 802.11h

Tabla III
COEFICIENTES DE CORRELACIÓN DE PEARSON

Variable 1	Variable 2	Coefficiente de Pearson
SNR	Distancia	-0,85
RSSI	Distancia	-0,54
RSSI	SNR	0,73
Jitter	SNR	-0,4
PLR	SNR	0,4
PLR	Jitter	0,92
Throughput	Errores RTP	-0,78
PLR	Errores RTP	0,89

En la Fig. 2 y Fig. 3 se muestran varias gráficas que demuestran cómo impacta el movimiento de los terminales WiFi en la provisión del servicio de VoIP. En ambas se muestra el instante de llegada de los paquetes al terminal en el eje horizontal y en el vertical se muestra el número de paquetes que llegan. La unidad de medida en el eje horizontal son: *horas:minutos:segundos*.

En la Fig. 2 se observa el resultado del experimento en el que en primer lugar los terminales están en cobertura. A partir del instante 20:04:55 se empiezan a alejar lo suficiente como para que empiece a llegar los paquetes a ráfagas (picos). Esta situación significa que el RSSI empieza a fluctuar y los terminales están en una zona de cobertura escasa (cuadro rojo). A continuación se vuelven a mover los terminales y se acercan de tal manera que los paquetes, después de un pico de llegada comienzan de nuevo a fluir con regularidad (cuadro azul). En este caso, la conexión de voz nunca se perdió definitivamente y únicamente se produjo una pérdida importante de calidad de la voz debido a la llegada fluctuante de los paquetes de voz.

En la Fig. 3 se muestra que en primer lugar los paquetes llegan con cierta regularidad porque los equipos están en cobertura. Entonces arrancamos una conversación de VoIP y se nota como los paquetes empiezan a llegar en gran cantidad (poco antes del 20:38:16). En ese momento empezamos a alejar los terminales y se observa como los paquetes empiezan a perderse por falta de cobertura (cuadrado rojo). Justo después ya no llega prácticamente ningún paquete y la conversación se pierde. A continuación se volvió a acercar los terminales (cuadro azul); pero sin embargo, la conversación de voz no se pudo recuperar, si en cambio la red *ad hoc* (por eso siguen llegando una cantidad ínfima de paquetes de sincronización).

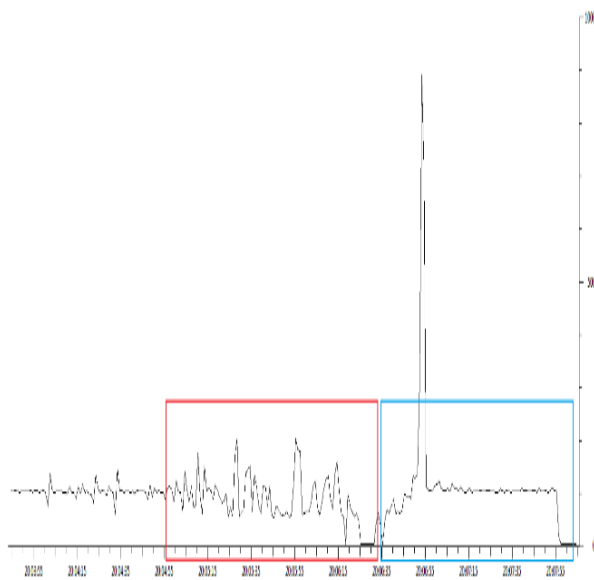


Fig. 2. Traza de tráfico de VoIP con Wireshark de paquetes recibidos para salir de cobertura.

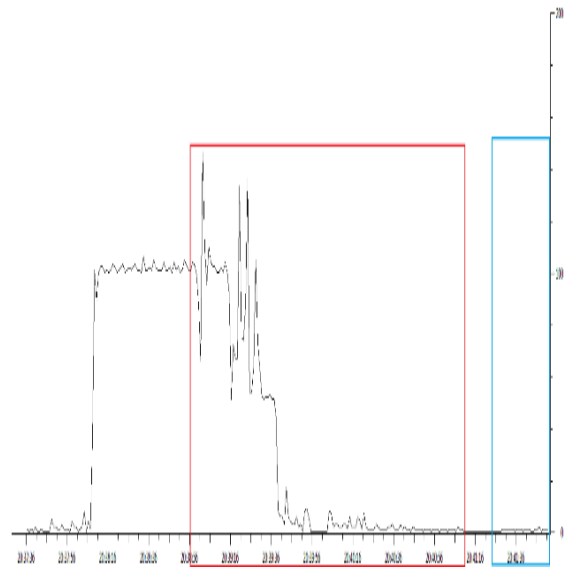


Fig. 3. Traza de tráfico de VoIP con Wireshark de paquetes recibidos para salir de cobertura y volviendo a entrar en cobertura.

B. Experimentos recientes con video streaming en redes inalámbricas

En este apartado presentamos algunas ideas básicas sobre la interrupción del servicio de video streaming en redes WiFi exponiendo un ejemplo en el que se puede apreciar la naturaleza de los problemas en este tipo de redes.

El procedimiento fue el siguiente: primero se establece la conexión entre los dos equipos de una red *ad hoc*. Una vez establecida dicha conexión se espera que el vídeo se vaya recibiendo de manera estable. En ese momento se hacen desconexiones manualmente del cliente utilizando una combinación de teclas que deshabilita la tarjeta de red interna del ordenador para a continuación volver a habilitarla. De esta forma vamos produciendo manualmente interrupciones de la comunicación para observar que influencia se produce en el envío y la recepción de paquetes de vídeo en la red inalámbrica. Para ello se instaló VideoLAN server en uno de los computadores y en el otro el VideoLAN para visualizar el vídeo. Con Wireshark capturando todos los paquetes de la interfaz inalámbrica se observó la traza de tráfico.

En la Fig. 4 se muestra el tráfico que se genera en el servidor y en el cliente con las desconexiones manuales que se provocan. En color azul se muestra el tráfico que se envía desde el servidor, mientras que en color rojo se muestra en que se envía desde el cliente justamente después de que intenta recuperar una sesión de video mediante *Real Time Streaming Protocol (RTSP)*. Se muestra que cada cierto tiempo, al provocar desconexiones no llega tráfico al cliente (situaciones en las que no hay sombreado azul ni verde). Entonces, el cliente trata de recuperar la sesión después de un tiempo (sombreado verde) enviando cierto tráfico para poder reestablecer la conexión con el servidor. Esto es necesario porque RTSP funciona sobre TCP y es obligado volver a iniciar la sesión si se pierde negociando de nuevo una sesión, justo en el caso que la desconexión haya sobrepasado cierto plazo de tiempo. Si no es el caso, entonces se puede recuperar el vídeo justamente por donde se dejó de recibir.

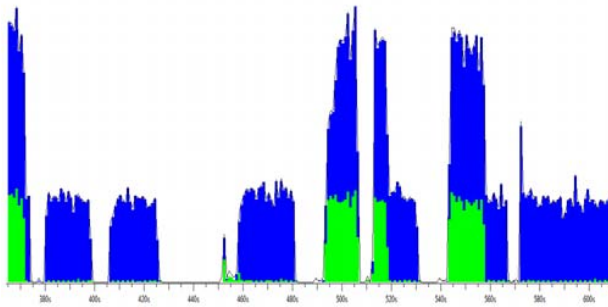


Fig. 4. Traza de tráfico de VoD con Wireshark de paquetes recibidos sin salir de cobertura.

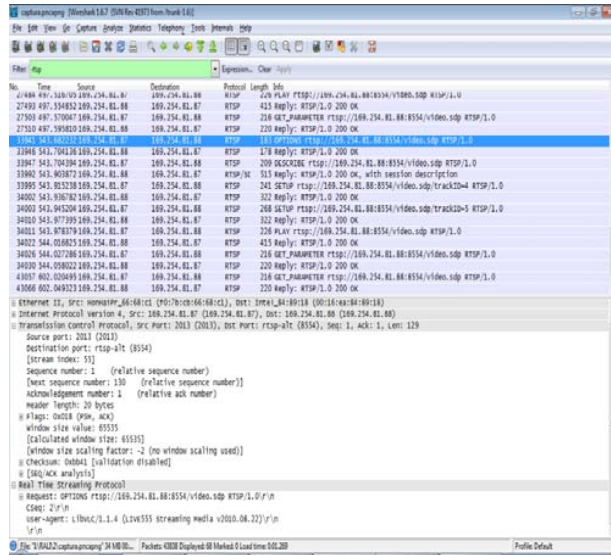


Fig. 5. Muestra de una re-negociación de una sesión RTSP desde el principio justo después de una desconexión larga.

En la Fig. 5 se muestra la situación en la que es necesario volver a reiniciar una sesión de video debido a que la desconexión duró un plazo largo de tiempo. Esto es, el TCP obliga a suspender la sesión de RTSP anterior y que se deba volver a negociar una nueva desde el principio.

En resumen, después de este experimento, podemos destacar lo siguiente:

- Cuando las desconexiones son cortas el vídeo vuelve a recibirse en cuanto se vuelve a conectar a la red *Ad hoc*.
- Si la desconexión es de mayor tiempo, no se recupera automáticamente porque el propio programa detiene la recepción pero en cuanto se quiere volver a reproducir el vídeo se recupera con normalidad.
- En los casos donde en los que el reproductor no se para, cuando vuelve a estar establecida la conexión ad hoc, el flujo de vídeo continúa por donde se está transmitiendo actualmente aunque intercala algunas tramas que quizás permanecían en el buffer.

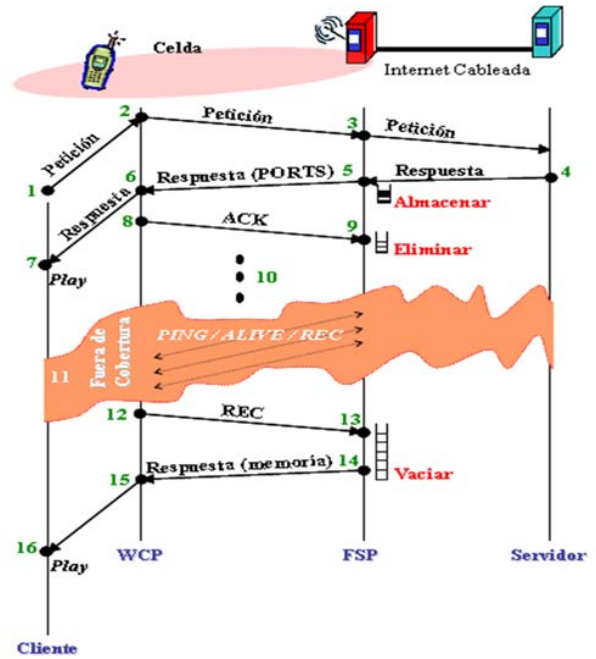


Fig. 6. Funcionamiento del mecanismo de nivel de aplicación para mitigar los problemas de interrupción del servicio de video Streaming en redes WiFi con teléfonos móviles.

- Ante una pausa en el cliente y reanudación, el vídeo se comporta de forma excelente continuando por donde se ha quedado y almacenando el resto en el buffer hasta que se llene.

C. Arquitectura de proxies para mitigar el problema de las desconexiones largas

En el año 2007 [9] hicimos la segunda publicación de un mecanismo de nivel de aplicación que permitía recuperar las sesiones de video streaming sin necesidad de volver a renegociar desde el principio una sesión nueva. En la Fig. 6 se muestra un esquema gráfico de cómo funciona este mecanismo adaptado para teléfonos móviles de la época.

El mecanismo era muy interesante porque, al ser los teléfonos móviles de aquella época muy simples, se debía aportar una solución minimalista muy ligera que fuera soportada por aquellos teléfonos. Ese mecanismo se ha modificado en varias ocasiones para adaptarlo a los nuevos tipos de terminales de cada momento. Sin embargo, sigue siendo válido debido a que todavía siguen existiendo teléfonos muy ligeros con *Firefox OS* que ejecutan servicios en la Nube.

Básicamente el mecanismo era el siguiente: entre los proxies (WCP y FSP) se intercambian mensajes de datos (usando RTP) que se confirman mediante ACK. Además se intercambian información sobre la conexión o desconexión del cliente a través de PING y ALIVE. Una vez reconectado, con la orden REC se informa del último paquete recibido

correctamente, para que el FSP proceda a enviar, a partir de él, los que tiene almacenados en su memoria de almacenamiento temporal. Aunque con la primitiva ACK se da indicación de los recibidos correctamente, es necesario indicar, en el momento de la reconexión cual fue el último, para evitar que se pierdan los ACK al estar ya el cliente fuera de cobertura.

Este esquema simple se fue modificando con los años para dotarlo de mayor capacidad de anticipación para mejorar la Calidad de experiencia de usuario (QoE, del inglés *Quality of Experience*) del usuario avisándole de posibles problemas futuros con las desconexiones [10]. A través de un sistema cross-layer que empleaban pre-alertas y alertas era posible anticipar información sobre el estado del canal utilizando básicamente el *Round Trip Time (RTT)*. A partir de este estudio fue posible iniciar el análisis de métodos más potentes usando agentes *Foundation for Intelligent Physical Agents (FIPA)* para implementar los proxies que tomaran decisiones inteligentes acerca de cuándo estimaban que se podía producir una interrupción del video. De esa forma se podía informar anticipadamente al usuario de cuando se podía dar dicha interrupción para que tomara acciones correctoras anticipadamente tal como se muestra en la Fig. 7.

Con el paso del tiempo estos mecanismos se han ido adaptando a WebRTC y el manejo de video en tiempo real entre navegadores directamente.

D. Experimentos recientes con Video Streaming usando navegadores Web directamente

En este experimento, utilizamos distintos tipos de vídeo almacenados directamente en *youtube* y localmente mediante *VideoLAN*. La idea es poder acceder desde un navegador usando HTTP para recibir el video. En este caso, se usó una red inalámbrica de un campus universitario y además se observaron distintos tipos de carga de la red WiFi (con más o menos usuarios conectados activos transmitiendo tráfico). La idea era obtener una imagen realista sobre el funcionamiento de video streaming almacenado teniendo en cuenta cargas reales de la red.

La idea era expandir el modelo de rendimiento presentado en para usar directamente navegadores caracterizando las interrupciones de servicio que se pueden dar de forma realista teniendo en cuenta nuestra arquitectura de *proxies* basadas en agentes y no teniéndola en cuenta.

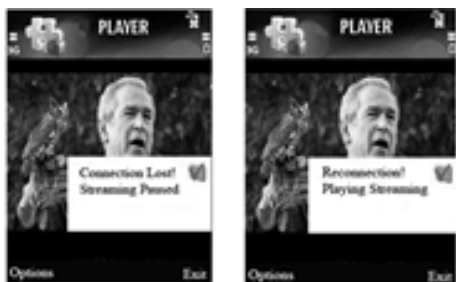


Fig. 7. Mensajes proactivos en el video para el usuario.

Tabla IV
DIFERENCIAS DE TIEMPO DE EJECUCIÓN CON Y SIN PROXIES PARA VIDEO STREAMING CON UNA SOLA INTERRUPCIÓN.

<i>Sin proxies</i>	<i>Con proxies</i>
718,52	675,66
721,63	674,76
723,14	676,26
722,90	676,96
724,69	676,66
735,69	678,36

Si no usamos la arquitectura de proxies, entonces cada vez que se produce una interrupción WiFi se ha de recomenzar de nuevo a recibir el vídeo desde el principio, con el consiguiente coste computacional y de la red. Esto era debido a que el servidor de video cierra la conexión, y la única forma de volver a ver el video es volver a empezar de nuevo desde el principio. Sin embargo, con nuestra arquitectura de proxies es posible no volver a empezar de nuevo desde el principio debido a que los proxies se encargan de mantener activa la sesión de video para que el servidor no la interrumpa.

Hicimos una gran cantidad de experimentos provocando una, dos... hasta 6 interrupciones en la conexión para observar construir un modelo de rendimiento aproximado que demostrara el rendimiento de la arquitectura de proxies. Todo ello observando los parámetros típicos de QoS. Estadísticamente se observaba que la reconexión con nuestra arquitectura de proxies tardaba lo mismo normalmente (tiempo que el navegador tardaba en reconectar con los proxies para recuperar la sesión de vídeo streaming). Sin embargo había casos en los que se tardaba más tiempo y estos casos aparecían aleatoriamente. Estos casos solían estar relacionados con la carga de usuarios en la red WiFi (podíamos acceder a la carga de usuarios en los PA de la red).

Así y todo siempre teníamos ventaja usando nuestra arquitectura de Proxies como se demuestra en la tabla IV. En esa tabla se muestran los tiempos de ejecución de 6 pruebas diferentes de la reproducción del vídeo con una sola interrupción medidos en ms.

Como se puede observar en los experimentos anteriores, con WiFi, no es sencillo proporcionar QoE debido a los problemas inherentes del canal inalámbrico.

A continuación presentamos brevemente las principales ideas de un método de estimación del RSSI como vía para adelantar información al nivel de aplicación (a nuestra arquitectura de proxies) que le permitan tomar acciones correctivas para paliar en la medida de lo posible las interrupciones de servicio multimedia. La ventaja de este método es que permite usarse con otros tipos de problemas de redes inalámbricas para mejorar su rendimiento.

E. Estimación eficiente del RSSI

Estimar eficientemente el RSSI con cierto tiempo de antelación es una vía para poder avisar al nivel de aplicación de posibles problemas que se pueden dar y prevenirla de que tome acciones correctivas que palien la pérdida de QoE. En

general este problema es irresoluble porque no hay manera de estimar el RSSI con antelación suficiente; pero en general cualquier aviso que se pueda hacer, aunque sólo sea con un segundo de antelación, podría suponer evitar interrupciones de servicio que mejorarían considerablemente la QoE.

Nosotros trabajamos la estimación del RSSI en [11]. En este trabajo pudimos observar ciertas propiedades del RSSI: básicamente que en ciertas condiciones solía tener un comportamiento casi lineal (realmente una componente lineal con una variabilidad mínima que se suele denominar desviación de la media). Gracias a este trabajo, pudimos demostrar que era posible anticipar información con bastante tiempo de antelación del RSSI en el caso que se moviera con velocidad constante. La idea era que podíamos inferir que nivel de RSSI tendría después de unos segundos si seguía moviéndose a cierta velocidad constante y mantenía el sentido de la trayectoria. Creamos un modelo de predicción del RSSI sencillo; pero muy efectivo y pudimos comprobarlo empíricamente. Este modelo se cumplía bajo ciertas restricciones: en campo abierto básicamente, y en ausencia de interferencias.

Este trabajo fue mejorado y presentado en una revista de alto impacto recientemente [12]. Sin embargo, la propuesta se hacía para Bluetooth y simplemente tomando ventanas de RSSI de tamaño fijo.

La aparición de un trabajo reciente [13] sobre RSSI ha demostrado que a lo largo del tiempo el RSSI muestra valores que tienden a retornar a medias móviles. Las medias móviles es un método estadístico que se utiliza para calcular la variabilidad de los valores de la bolsa, en Economía. Curiosamente, estos autores demostraron que con el RSSI se podía aplicar estos modelos econométricos.

Haciendo el análisis de este modelo, nosotros en [14] hemos demostrado que se puede simplificar dicho modelo para no tener que incluir una componente aleatoria compleja que hace que dicho método sea de difícil aplicabilidad en la práctica. Además, [13] no estaba probado empíricamente para distintos tipos de terminales, sistemas operativos y situaciones realistas, como nosotros hicimos.

Posteriormente, demostramos que los resultados de este modelo se podían mejorar si incluíamos ventanas de RSSI de tamaño variable [15]. La observación fue sencilla: algunos valores de RSSI no llegaban al terminal en el instante de tiempo estipulado (*beacon interval*), sino que podía llegar después de ese tiempo una cantidad aleatoria que depende, entre otras cosas, de la cantidad de tráfico cursado en el canal. Teniendo en cuenta este hecho y aplicando nuestro método original dentro de una ventana, se logra una estimación del RSSI que mejora muchísimo a los métodos previos. Esto es posible porque en realidad nuestro método original estimaba la tendencia del RSSI al retorno a las medias móviles. Combinado esto con la construcción de un algoritmo que construye ventanas de tamaño variable adaptadas al estado del canal podemos estimar el RSSI eficientemente.

Con esta estimación, se puede avisar con tiempo de antelación suficiente al proxy-cliente multimedia de nuestra arquitectura para que tome acciones correctivas que mitiguen los efectos adversos de las interrupciones multimedia que empeoran la QoE.

IV. CONCLUSIONES

En este artículo hemos presentado la línea de trabajo en QoE para servicios multimedia en redes inalámbricas, en la que trabaja nuestro grupo de investigación.

En resumen hemos intentado poner de manifiesto como a lo largo de estos años hemos ido haciendo experimentación práctica con servicios multimedia (videostreaming y VoIP básicamente) y obteniendo diversos resultados que ponen de manifiesto la problemática de las interrupciones del servicio debido a múltiples factores: el medio físico (radio) de transmisión es poco fiable, el MAC en WiFi sigue teniendo muchos problemas sin resolver (aunque se han propuesto una cantidad enorme de soluciones que individualmente no funcionan correctamente), el TCP no ajusta bien sus ventanas de congestión y control de flujo porque no distingue errores de transmisión de desconexiones del canal radio... Y a nivel de aplicación, el software actual no está preparado para soportar eficazmente las desconexiones.

Nosotros hemos planteado el uso de una arquitectura de proxies que permiten controlar e incluso intentan anticipar (cuando es posible) las posibles interrupciones de servicio. Esta arquitectura ha demostrado ser eficiente en distintos casos y en estos momentos tenemos previsto aplicar las nuevas tecnologías de diseño de redes: Software Defined Networking (SDN) y Network Function Virtualization (NFV). Creemos que nuestros proxies son lo suficientemente sencillos de programar. Por ello, se podría intentar aplicar NFV para virtualizar el funcionamiento de las primitivas de nivel de enlace al objeto de que esas funciones propagasen valores a los niveles superiores de la arquitectura de red (Internet), intentando mejorar el rendimiento del software multimedia de nivel de aplicación. En cualquier caso, esto es una idea de prospectiva que debe ser evaluada con más detalle.

AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación parcial de la ULPGC.

REFERENCIAS

- [1] M. Antonio, "Teletrófono | IDIS (no date). Available at: <http://proyectoidis.org/antonio-meucci-teletrono/> (Accessed: 17 May 2015)," 2015.
- [2] A. Suarez, G. Blacio, E. Macias, and B. Ramos, "Arquitectura de VoIP en la nube con interfaz web, verde, open source y de bajo coste," *Actas de JITEL 2013.*, 2013.
- [3] "Precusores de la Televisión. Entradas sobre Paul Nipkow en Historia de los medios (no date) Available at: <https://lahistoriadelosmedios.wordpress.com/tag/paul-nipkow/> (Accessed: 17 May 2015)," 2015.
- [4] J. Merodio, "Qué es el Cloud TV y su Importancia en la Social TV. Available at: <http://www.juanmerodio.com/2013/que-es-el-cloud-tv-y-su-importancia-en-la-social-tv/> (Accessed: 18 May 2015)," May. 2013.
- [5] K. Atalah, "Contributions to effective protocol design to mitigate continuous multimedia services short disruptions in WiFi networks, Ph.D. Thesis (A. Suarez & Elsa Macias supervisors)," 2011.
- [6] Google, "Project Fi," [Online; accessed 28-April-2015] <<https://fi.google.com/about/network/>>, 2015.
- [7] J.M. Alvaro Suárez Elsa Macías and V. Sunderam, "Using OLSR for Streaming Video in 802.11 Ad Hoc Networks to Save Bandwidth," *IAENG International Journal of Computer Science*, 2007.
- [8] P. Demestichas, A. Georgakopoulos, D. Karvounas, K. Tsagkaris, V. Stavroulaki, J. Lu, C. Xiong, and J. Yao, "5G on the Horizon: Key Challenges for the Radio-Access Network," *Vehicular Technology Magazine, IEEE*, vol. 8, 2013, pp. 47-53.

- [9] E.M. Alvaro La-Menza M. Macias-Lopez Suarez and V. Sunderam, "Automatic resumption of streaming sessions over WiFi using JADE," *IAENG International Journal of Computer Science*, vol. 33, 2007, pp. 92–100.
- [10] M. Elsa and A. Suarez, "Fourth-Generation Wireless Networks: Applications and Innovations," A.M. Sasan Adibi and M. Tofighbakhsh, eds., IGI Global, 2010, pp. 494–525.
- [11] A. Suárez, K.A. Elbatsh, and E. Macias, "Gradient rssi filter and predictor for wireless network algorithms and protocols," *Network Protocols and Algorithms*, vol. 2, 2010, pp. 1–26.
- [12] F. Subhan, S. Ahmed, K. Ashraf, and M. Imran, "Extended Gradient RSSI Predictor and Filter for Signal Prediction and Filtering in Communication Holes," *Wireless Personal Communications*, 2015, pp. 1–18.
- [13] E. Chin, D. Chieng, V. Teh, M. Natkaniec, K. Loziak, and J. Gozdecki, "Wireless link prediction and triggering using modified Ornstein–Uhlenbeck jump diffusion process," *Wireless Networks*, vol. 20, 2014, pp. 379–396.
- [14] A. Suárez, J.A. Santana, E.M. Macias-Lopez, V.E. Mena, J.M. Canino, and D. Marrero, "RSSI Prediction in WiFi Considering Realistic Heterogeneous Restrictions," *Network Protocols and Algorithms*, vol. 6, 2014, pp. 19–40.
- [15] J.A. Santana, E. Macias-Lopez, V.E. Mena, D. Marrero, and A. Suárez, "Estimación Eficiente del RSSI en Redes WiFi para Servicios Inalámbricos Futuros," *Actas JITEL 2015. En revisión.*, 2015.

Impacto de Parámetros de QoS en Aspectos de QoE: Análisis desde el Punto de Vista de la Sincronización Multimedia

Mario Montagud^{1,2}, Fernando Boronat², Marc Martínez², Jordi Belda², Pablo Cesar¹

¹Centrum Wiskunde & Informatica (CWI), Amsterdam

²Universitat Politècnica de València (UPV) – Campus de Gandia

montagud@cwi.nl; {fboronat@dcom., marmarc3@upvnet., jorbelva@epsg.}.upv.es; P.S.Cesar@cwi.nl

Resumen- La sincronización multimedia ha sido un área de investigación clave desde los inicios de los sistemas multimedia. En este artículo se ofrecen una visión general y un análisis sobre el impacto de varios parámetros de QoS en diferentes aspectos de la QoE, desde el punto de vista de la sincronización multimedia. En primer lugar, se presentan los diferentes tipos de sincronización multimedia y su relevancia para garantizar una QoE satisfactoria. En segundo lugar, se muestra que la magnitud de los retardos y su variabilidad en las redes actuales es bastante superior a los umbrales permisibles por los usuarios en diferentes tipos y ejemplos de sincronización multimedia, reflejando así su necesidad. En tercer lugar, se describe el impacto del ancho de banda y la tasa de pérdidas sobre la sincronización multimedia. Por último, se argumenta la influencia del uso de diferentes alternativas para conseguir la sincronización multimedia sobre varios factores de QoS y de QoE.

Palabras Clave- IDMS, QoS, QoE, Sincronización Híbrida, Sincronización Multimedia

I. INTRODUCCIÓN

Las redes actuales de conmutación de paquetes no ofrecen garantías ni una respuesta estable y óptima para la distribución de contenidos multimedia. Esto se convierte en un desafío, más aún cuando existen requisitos estrictos en cuanto a retardo, interactividad, tasa de pérdidas, ancho de banda y/o escalabilidad. En los últimos años se ha estado investigando con el objetivo de mejorar las prestaciones tanto de las tecnologías de red como de los sistemas finales y de interconexión, para poder ofrecer servicios multimedia sobre redes IP de manera satisfactoria. Algunos de los avances más relevantes que se pueden destacar son: redes de mayor alcance y capacidad, mayores velocidades y capacidades de computación, códecs más avanzados y eficientes, nuevas tecnologías de transmisión adaptativas (p.ej., Hypertext Transfer Protocol o HTTP streaming, Datagram Congestion Control Protocol o DCCP...), políticas de admisión, reserva de recursos y asignación de prioridades, mecanismos de encaminamiento avanzados, sistemas de replicación de contenido y caching, etc.

En las comunicaciones telemáticas y, especialmente, en los servicios multimedia interactivos, los parámetros de calidad de servicio (Quality of Service o QoS), como pueden ser la disponibilidad, el retardo y su variabilidad (ya sea para un mismo flujo de datos, es decir el jitter, como entre flujos y entre receptores), la tasa de pérdidas y el ancho de banda disponible, tienen un impacto muy claro sobre varios aspectos de la calidad de experiencia (Quality of Experience

o QoE) percibida por los usuarios, como son la satisfacción, la perceptibilidad, la molestia, la inmervisidad, el *engagement*, la conectividad virtual (*networked togetherness*), la igualdad de condiciones (*fairness*), etc. Así pues, es necesario optimizar los aspectos tecnológicos, estrechamente relacionados con la QoS, así como los aspectos operacionales, con tal de brindar a los usuarios la mejor QoE posible, pues ello va a determinar el éxito de los servicios multimedia existentes y de los que se puedan desarrollar en el futuro.

En este artículo se ofrecen una visión general y un análisis sobre la influencia de varios parámetros de QoS, centrándose especialmente en los retardos y su variabilidad, en diversos aspectos de QoE, desde el punto de vista de la sincronización multimedia. Se van a presentar los diferentes tipos de sincronización multimedia, destacando algunos de los casos de uso más relevantes, y se van a identificar los desafíos tecnológicos, ligados a parámetros de QoS, que pueden frustrar la QoE percibida por los usuarios. En particular, se van a identificar umbrales tolerables para diferentes tipos y casos de uso de sincronización multimedia, así como también se van a describir algunas implicaciones con respecto al uso de varias alternativas para conseguir la sincronización multimedia. Asimismo, a lo largo del artículo se va a identificar la necesidad de seguir investigando en ciertos aspectos, tanto para determinar los umbrales permisibles como para diseñar soluciones que mejoren la QoE.

Este artículo se ha estructurado de la siguiente manera. En la Sección II se presentan los diferentes tipos de sincronización multimedia, en sus diferentes variantes, incluyendo algunos ejemplos relevantes. En la Sección III se enumeran los diferentes factores que pueden contribuir al incremento de los retardos y su variabilidad en las comunicaciones multimedia. Asimismo, se recogen resultados de medidas de retardos en diferentes escenarios reales. En la Sección IV se presentan los umbrales tolerables por la percepción humana en cuanto a errores (esto es, asincronías) en diferentes tipos y ejemplos de sincronización multimedia. En la Sección V se presenta brevemente la influencia del esquema de control para la sincronización utilizado en varios factores de QoS y aspectos de QoE. En la Sección VI se discute el impacto del uso de varias estrategias para la selección de un flujo o componente multimedia como la referencia temporal para la sincronización. El artículo finaliza con las conclusiones en la Sección VII.

II. TIPOS DE SINCRONIZACIÓN MULTIMEDIA

Se pueden distinguir varios tipos y variantes de sincronización multimedia (véase la Fig. 1), en función del número de componentes multimedia, de emisores, flujos y receptores involucrados. Primero, la sincronización intra-media (también conocida como sincronización intra-flujo) es necesaria para mantener las relaciones temporales originales entre las unidades de datos, Media Units o MUs (p.ej., tramas de video, muestras de voz...), que forman parte de cada componente multimedia (p.ej., video, audio...). Estas relaciones temporales se suelen especificar en los procesos de captura o emisión y pueden verse alteradas durante la cadena de distribución multimedia extremo-a-extremo, debido principalmente al jitter y las pérdidas. Es por ello que se necesitan mecanismos precisos y adaptativos para mantenerlas (y restablecerlas, si es necesario) de manera que el contenido multimedia se reproduzca de acuerdo a los patrones temporales apropiados. Típicamente, se utilizan buffers de reproducción con tal de compensar el efecto del jitter y de las pérdidas. Si no se incluyen mecanismos de sincronización intra-media, existirán distorsiones temporales en la reproducción individual de cada componente multimedia (p.ej., chasquidos en el audio, pausas y acelerones en el video...), y esto puede llegar a ser muy molesto para los usuarios (es decir, se corresponderá con una mala QoE).

Segundo, la sincronización inter-media (también conocida como sincronización inter-flujo) se ocupa de preservar las dependencias temporales entre las MUs de diferentes componentes multimedia (p.ej., la sincronización labial o *lip-sync*). La falta de sincronización inter-media puede causar confusión a los usuarios (p.ej., comentarios de audio o subtítulos que no se corresponden con las escenas que se visualizan), por lo que puede llegar a ser irritante (mala QoE). Se pueden distinguir dos situaciones diferentes en cuanto a la sincronización inter-media. La primera de ellas es cuando los diferentes componentes multimedia se multiplexan en un único flujo de datos, tal y como se puede observar en la Fig.2.a.1 (en este caso los términos sincronización inter-media e inter-flujo no se pueden utilizar como sinónimos, al existir un único flujo). La segunda de ellas es cuando cada componente multimedia se envía en un flujo de datos independiente, tal y como se puede observar en la Fig.2.a.2. En el primer caso, conseguir la sincronización inter-media es más fácil, ya que los componentes multimedia están contenidos en el mismo flujo y, por tanto, han sufrido prácticamente el mismo retardo, así como también es más fácil identificarlos y enlazarlos. Sin embargo, el segundo caso otorga más flexibilidad y eficiencia con respecto al uso de ancho de banda. Por ejemplo, un usuario podría recibir sólo los componentes multimedia que le interesan, o cesar temporalmente la recepción de algún componente multimedia (considerado menos prioritario) en situaciones de congestión. La primera alternativa se suele utilizar en tecnologías MPEG (p.ej., Moving Picture Experts Group 2 Transport Stream o MPEG2-TS, Dynamic Adaptive Streaming over HTTP o DASH...), mientras que la segunda es más común cuando se utilizan los protocolos Real-Time Transport Protocol / RTP Control Protocol (RTP/RTCP) [1] para la entrega de contenidos multimedia.

Un sub-tipo específico de la sincronización inter-media es la sincronización inter-emisor (también conocida como sincronización multi-fuente), y se ocupa de sincronizar la reproducción de varios componentes multimedia enviados por diferentes emisores o fuentes (véase la Fig. 2.b). También puede suceder que los diferentes componentes multimedia se envíen mediante diferentes protocolos o incluso a través de diferentes redes, de la misma o diferente naturaleza. En este último caso, se suele llamar sincronización híbrida (*hybrid synchronization*) [2]. Un ejemplo muy relevante en el mercado de TV actual es la sincronización entre flujos broadcast y broadband, como pueden ser flujos DVB (Digital Video Broadcasting) y flujos obtenidos a través de HTTP streaming (p.ej., DASH, HTTP Live Streaming o HLS...). Esta es una de las posibilidades que puede ofrecer el estándar HbbTV (Hybrid Broadcast Broadband TV) [3]. A través de un uso simultáneo y coordinado de las redes broadband y broadcast, la sincronización híbrida permite ofrecer servicios multimedia enriquecidos y personalizados a los usuarios, y por tanto una mejor QoE. El contenido ofrecido por la red broadcast se puede complementar con contenido adicional relacionado ofrecido por redes broadband. Escenarios típicos son la selección de videos multi-cámara, la selección del canal de audio o emisora de radio favorito/a, el acceso a contenido web complementario, todo ello a través de redes broadband. Esto permite una mayor flexibilidad, un mayor abanico de posibilidades y un ahorro significativo de costes, debido al alto precio del ancho de banda broadcast (licencias para el uso del espectro).

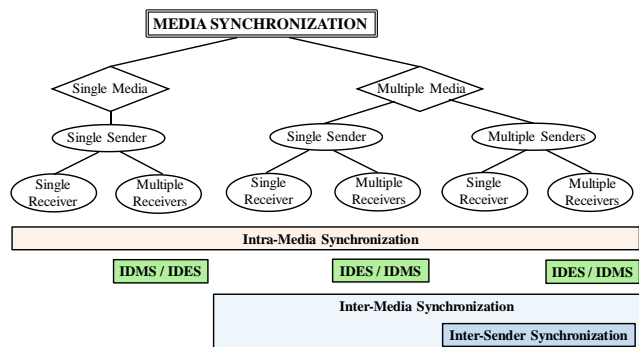


Fig. 1. Visión general del área de investigación relativa a la sincronización multimedia.

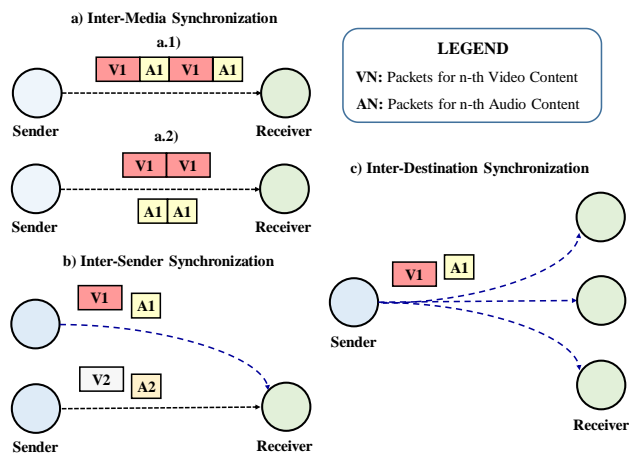


Fig. 2. Diferentes tipos y variantes de sincronización multimedia.

Además de la sincronización de componentes multimedia en cada receptor específico (p.ej., TV, PC...), también es necesaria la sincronización simultánea de los procesos de reproducción de componentes multimedia en diferentes receptores o dispositivos (véase la Fig.2.c). Esto se conoce como sincronización inter-dispositivo (Inter Device Synchronization o IDES), y también como sincronización inter-destinatario (Inter-Destination Synchronization o IDMS), especialmente cuando los dispositivos están separados geográficamente (p.ej., en Social TV, servicios de multi-conferencia, juegos en red multi-jugador...) [4]. La sincronización IDMS es muy relevante para posibilitar una gran variedad de experiencias multimedia compartidas entre usuarios remotos (hasta 20 casos de uso se describen en [5]). Su inexistencia puede afectar negativamente a la QoE, por ejemplo, causando interacciones incoherentes o malentendidos en las comunicaciones en grupo, desigualdad de condiciones en escenarios de ámbito competitivo (p.ej., juegos en red), lo que puede ser muy desesperante para los usuarios. El ejemplo típico es el de un usuario viendo un partido de fútbol online con sus amigos de forma remota, y percibir un gol a través de los canales de chat, o de los gritos de sus vecinos, antes de ver el gol en su propia pantalla.

III. VARIABILIDAD DE RETARDOS EN LAS COMUNICACIONES

Se ha comentado previamente que los retardos y su variabilidad, ya sea para un mismo flujo (jitter), entre diferentes flujos o entre diferentes receptores, son los factores más determinantes por los que se van a necesitar soluciones de sincronización multimedia. En primer lugar, en esta sección se presentan los diferentes factores a lo largo de la cadena de distribución multimedia que contribuyen al incremento de la magnitud de los retardos y de su variabilidad. En segundo lugar, se proporcionan datos de medidas realizadas en estudios de investigación recientes que reflejan la magnitud de los retardos extremo-a-extremo y de su variabilidad en redes de distribución actuales.

Los retardos y su variabilidad no son un problema serio cuando se consume contenido multimedia almacenado de manera aislada, pero sí se convierten en una barrera cuando se consume contenido multimedia en vivo y/o diferentes usuarios están interactuando de manera remota en el contexto de un consumo conjunto de contenido multimedia.

A. Factores que Contribuyen a los Retardos y a su Variabilidad

Cuando se distribuye contenido multimedia por Internet, cada uno de los componentes multimedia involucrados (p.ej., audio, video...) se estructura, durante el proceso de captura o lectura, en una secuencia de MUs (p.ej., tramas de video o muestras de voz) que, a su vez, se encapsularán en uno o varios paquetes de datos para su transmisión por las redes IP. Cada uno de estos paquetes puede seguir la misma o diferente ruta que los demás paquetes del mismo o de diferentes componentes multimedia, así como también puede verse afectado de manera diferente por el jitter.

Tal y como se puede apreciar en la Fig. 3, a lo largo de la cadena de distribución multimedia extremo-a-extremo varios factores pueden contribuir al incremento de la magnitud de los retardos y de su variabilidad [4,6], pudiendo ambos oscilar desde unos pocos milisegundos hasta varios segundos.

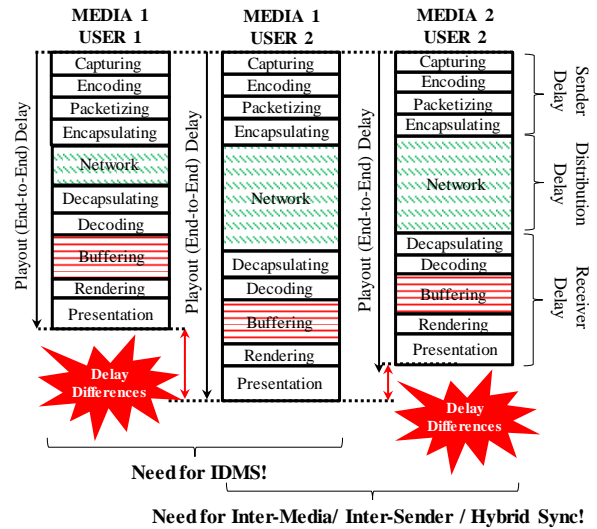


Fig. 3. Variabilidad de retardos: Necesidad de sincronización multimedia.

En el lado del emisor, procesos como la captura, codificación, cifrado, buffering, paquetización y encapsulamiento, entre otros, pueden contribuir al incremento de los retardos y de su variabilidad. Por ejemplo, el estudio en [7] demuestra que los tiempos de captura y codificación pueden diferir significativamente, en el orden de decenas de milisegundos, para diferentes flujos de audio y video, así como para diferentes procesadores, dependiendo también de su carga puntual de proceso.

En la red de distribución, varios factores pueden afectar al incremento de los retardos, así como de su variabilidad, como son los procesos de propagación y transmisión de paquetes, tanto a través de los enlaces de comunicaciones como de los dispositivos intermedios y finales involucrados, los procesos de buffering y control en los routers intermedios (p.ej., decisiones de encaminamiento, políticas de QoS...), así como otros procesos avanzados como son la fragmentación y reensamblado de paquetes, trans-codificación, conversión de formatos, etc. Además, la carga instantánea de la red va a influir significativamente en la magnitud de estos retardos y de su variabilidad, tanto para un mismo flujo multimedia (es decir, el jitter), como entre diferentes flujos multimedia recibidos por el mismo o diferentes dispositivos.

En el lado del receptor, varios factores también contribuirán al incremento de los retardos y de su variabilidad, como son los procesos de buffering, descifrado, decodificación, de-paquetización o procesado en los diferentes niveles de la pila de protocolos, técnicas de detección y corrección de errores, retardos de entrega (*rendering delay*), etc. Quizás, de entre todos ellos, los retardos de buffering y decodificación sean los más relevantes. Estos retardos pueden fluctuar con respecto al tiempo, así como también pueden diferir entre diferentes receptores, debido a las diferentes capacidades de procesado, a la carga instantánea de CPU (Central Processing Unit), a la respuesta en tiempo real del sistema operativo en cuestión, al software y hardware disponible, así como a la carga instantánea de la red (pues determinará la tasa a la que se recibe el contenido multimedia).

Otro factor adicional a tener en cuenta cuando se utilizan TV digitales es el retardo de display (*display lag*), que es la diferencia entre el instante en el que una imagen (procedente

de una señal de video) se entrega a un display y el instante en el que dicha imagen se visualiza. Este retardo se origina debido principalmente a rutinas de procesamiento de la imagen, como es el escalado. El *display lag* puede originar una asincronía entre la señal de video y la de audio, afectando, por tanto, a la sincronización inter-media (si no se compensa de algún modo). El *display lag* en TVs de alta definición (High Definition or HD TV) puede oscilar entre 30 y 90ms, dependiendo del modelo y del tipo de señal de video de entrada [8]. En este contexto, el estudio en [9] analizó el impacto de diferentes modos de TV (concretamente, el modo "Dinámico" y el modo "Juego") sobre el retardo de video y su variabilidad. Para la TV bajo test, se observó que el retardo en modo "Juego" fue casi 90ms menor y un poco menos variable (a costa de una ligera degradación de calidad, según la documentación de la TV, aunque fue inapreciable para los autores de dicho estudio). En concreto, en el escenario bajo test, para el modo "Juego", el retardo extremo-a-extremo medio fue de 357ms (con una desviación estándar de 20ms), mientras que para el modo "Dinámico", fue de 442ms (con una desviación estándar de 36ms). Por tanto, las características y configuración de los dispositivos finales pueden tener un impacto significativo sobre el retardo y su variabilidad.

Otro factor muy importante que puede contribuir a las diferencias de retardos, tanto para un mismo flujo como entre varios flujos, y tanto para un mismo receptor como entre diferentes receptores, son las imperfecciones de los relojes, pues van a acarrear desviaciones (lineales y/o aleatorias) en las tasas de generación, transmisión y reproducción de los contenidos multimedia.

Un análisis más exhaustivo sobre la contribución e impacto de cada uno de los factores mencionados en esta sección sobre la magnitud de los retardos y de su variabilidad se puede consultar en [6].

B. Magnitud de los Retardos en Redes Actuales

Después de presentar los diferentes factores que contribuyen a los retardos y a su variabilidad, esta sección proporciona datos sobre su magnitud en escenarios reales, tanto para un mismo flujo, como para diferentes flujos enviados a un mismo o diferentes receptores, de acuerdo a medidas realizadas en estudios relacionados.

En primer lugar, el estándar International Telecommunication Union (ITU) G.1050 [10] indica que los valores de retardos típicos en Internet pueden oscilar entre 20 y 500ms, mientras que los valores de jitter pueden oscilar entre 0 y 500ms. Por otro lado, el estándar ITU-T (ITU Telecommunication Standardization Sector) G.114 [11] indica que se requieren retardos inferiores a 150ms para aplicaciones de conferencia sobre Internet, mientras que retardos superiores a 400ms son típicamente inaceptables.

En [12], se indica que los retardos extremo-a-extremo cuando se utilizan herramientas de videoconferencia convencionales en un entorno LAN (Local Area Network) pueden oscilar entre 99ms, para Google Talk (Gtalk), y 312ms (con una desviación estándar de 67ms), para Skype. Asimismo, retardos extremo-a-extremo en torno a 350ms (también con una desviación estándar de 67ms) se obtuvieron cuando se utilizó una herramienta de videoconferencia ad-hoc en un escenario de red distribuido entre Bélgica y el Reino Unido.

En [13], se realizaron medidas de retardos en diferentes escenarios de red. A partir de estas medidas, se hizo un análisis de los mejores y peores casos para diferentes factores de retardo a lo largo de la cadena de distribución multimedia, cuyo resumen se puede consultar en la Tabla I. Se puede apreciar que los procesos de codificación, transmisión y buffering son los que más retardos introducen. Además, se comprobó en este estudio que las diferencias de retardo extremo-a-extremo entre diferentes receptores en un escenario IPTV pueden llegar a ser superiores a 6s (con retardos que oscilan entre 250ms y 6500ms).

El trabajo de investigación en [8] proporciona medidas sobre diferencias entre tiempos de reproducción cuando se recibe el mismo contenido multimedia utilizando diferentes tecnologías de TV, como son diferentes variantes del estándar DVB, redes de cable analógicas, IPTV y TV vía Web. Se indica que las diferencias de retardos entre tecnologías broadcast pueden llegar a ser en torno a 5s, incluso hasta 8s si se comparan con tecnologías de TV vía Web.

De manera similar, el estudio en [14] proporciona medidas sobre la magnitud de las diferencias de retardos para diferentes tecnologías de TV. Se indica que las diferencias de retardos entre tecnologías TV broadcast en un entorno nacional (Holanda) pueden llegar a ser del orden de 5s, mientras que en un entorno internacional (entre Holanda y el Reino Unido) pueden llegar a ser del orden de 6s. Asimismo, estas medidas reflejan que las tecnologías broadcast analógicas tienen un retardo significativamente menor que las digitales y que, en general, las señales HD tienen un retardo ligeramente superior a las de definición estándar (Standard Definition o SD). Además, se comprobó que las diferencias de retardos entre tecnologías de TV broadcast y tecnologías de TV basadas en Web (broadband), como DASH, pueden llegar a ser superiores a 1 minuto. Por último, en [14] también se percibieron diferencias de retardos entre receptores muy significativas cuando se utiliza exactamente el mismo dispositivo de consumo, la misma tecnología de TV y la misma configuración (p.ej., mismo tipo de suscripción y misma calidad de señal) en cada uno de ellos. Sin embargo, no se proporcionan magnitudes para estas diferencias, debido a la falta de resultados concluyentes, ya que, según los autores, no se disponía de suficientes medidas para diferentes combinaciones y números de usuarios distribuidos geográficamente, utilizando diferentes tipos de tecnologías y configuraciones.

Tabla I
FACTORES QUE CONTRIBUYEN AL RETARDO Y A SU VARIABILIDAD [13]

Factores		Rango de Retardos (ms)
Emisión	Captura de Video	17 - 40
	Codificación de Video	50 - 2000
	Cifrado	0 - 50
	Protección frente a errores	0 - 100
	Buffer de transmisión	50 - 500
Red	Transmisión de subida	10 - 300
	Trans-codificación	0 - 2000
	Transmisión de bajada	10 - 300
Recepción	Jitter buffer	50 - 500
	Protección frente a errores	0 - 100
	Descifrado	0 - 50
	Decodificación de Video	50 - 500
	Display buffer	0 - 50
TOTAL		250 - 6500

IV. UMBRALES TOLERABLES PARA LA PERCEPCIÓN HUMANA

Conocida ya la magnitud de los retardos y de su variabilidad en escenarios reales, en esta sección se presentan los umbrales tolerables en cuando a diferencias de retardos en diferentes tipos y ejemplos de sincronización multimedia, basándose en las conclusiones de trabajos de investigación relacionados. Asimismo, se va a argumentar brevemente y de manera cualitativa el impacto de otros parámetros de QoS, como son el ancho de banda y la tasa de pérdidas, sobre la percepción humana (QoE), desde el punto de vista de la sincronización multimedia.

A. Niveles de asincronías permisibles

Desde los inicios de las comunicaciones multimedia, la problemática del jitter ha sido palpable, pues repercute significativamente en los niveles de fidelidad e inteligibilidad de las señales multimedia, incluso en entornos locales. De hecho, ya se pueden encontrar en la década de los 70 estudios que analizaron el impacto del jitter en la sincronización intra-media. Por ejemplo, en [15] se defiende que la máxima asincronía o diferencia de retardos, también referida en algunas publicaciones como desviación (*skew*) o error, para la sincronización intra-media en audio de alta calidad es de unos 200ns.

Asimismo, con la introducción de los primeros servicios de contenido bajo demanda, CoD (Content on Demand), y de audio/video conferencia, la sincronización inter-media (principalmente *lip-sync*) se identificó como un problema a solucionar, por lo que varios estudios han tratado de determinar los umbrales tolerables para la percepción humana (QoE) en estos casos. Por ejemplo, los estudios subjetivos realizados en [16] reflejan que asincronías inferiores a 80ms para *lip-sync* son indetectables, mientras que asincronías superiores a 160ms son inaceptables. En [17], se indica que asincronías entre -80ms (audio atrasado con respecto al video) y +80ms (audio adelantado con respecto al video) son perceptibles, pero tolerables para la mayoría de los usuarios, mientras que niveles de asincronía superiores a -240ms y +160ms son intolerables. En este trabajo, además, se clasifican diferentes niveles de QoS en cuanto a límites de asincronía, dependiendo del tipo de información multimedia, sus parámetros y la aplicación específica, oscilando entre requisitos de sincronización muy estrictos para audio/audio (+/-11µs) hasta requisitos menos estrictos para audio/puntero (-500ms to 750ms).

De manera similar, en [18] se clasifican los umbrales para *lip-sync* en 3 rangos: imperceptibilidad (desde -95ms hasta +25ms), perceptibilidad (desde -125 hasta +45ms) y aceptabilidad (desde -185 hasta +90ms). Estos umbrales se ilustran en la Fig.4. El eje vertical de esta figura representa una diferencia constante entre el umbral perceptible y el aceptable (45ms para audio adelantado y 60ms para audio atrasado). El área roja representa los rangos intolerables para audio adelantado y atrasado, mientras que el área verde representa los límites de imperceptibilidad.

En [19], se recomiendan niveles más estrictos para *lip-sync*, estableciendo los niveles aceptables entre +15ms para audio adelantado y -45ms para audio atrasado.

Normalmente, en los estudios sobre perceptibilidad relacionados con *lip-sync* se concluye que los usuarios son más sensibles a asincronías cuando el audio está adelantado

con respecto al vídeo, que al revés. Esto se puede explicar debido a que la velocidad de la luz es mucho mayor que la velocidad del sonido y, por tanto, en condiciones normales (es decir, asumiendo la misma distancia a la pantalla y a los altavoces), la señal de audio llega más tarde a los usuarios que la señal de vídeo [20].

En [21] se analizaron los niveles de asincronía permisibles entre varias señales acústicas (estéreo) para un array de micrófonos. Se concluyó que una asincronía de 17ms entre las señales de audio es perceptible y que es preferible que la asincronía no supere los 11ms.

En [22] se evaluó la calidad de sincronización inter-media entre los componentes de video en videos 3D estereoscópicos, con una tasa de 25 tramas por segundo (*fps, frames per seconds*) y con una resolución espacial de 1920x1080. Se indica que una asincronía inferior a 80ms garantiza una buena calidad visual 3D, mientras que la QoE es inaceptable si la asincronía es superior a 200ms.

En [23] se evaluaron subjetivamente los niveles de asincronía aceptables entre un flujo de video y un flujo de datos multimedia provenientes de dispositivos que incluyen información sobre el tacto (*haptics*). Se demuestra que asincronías en torno a 40-80ms son apenas perceptibles, mientras que asincronías superiores a 300ms son molestas.

En [24] se realizaron estudios subjetivos para evaluar los niveles de asincronía tolerables entre aromas y contenido audiovisual, asumiendo una *lip-sync* perfecta. La conclusión fue que los umbrales de sincronización son de 30s cuando los aromas se presentan antes que el contenido audiovisual y de 20s cuando se presentan después.

Combinando los dos estudios anteriores, en [25] se investigaron los límites permisibles para la sincronización entre aromas y haptic, estableciendo los umbrales permisibles entre 1s y 3s.

Con respecto a la sincronización de subtítulos con contenido audiovisual, no se dispone de conocimiento sobre la existencia de estudios que hayan tratado de identificar los umbrales de sincronización permisibles. Esto podría ser debido a que los subtítulos normalmente se enviaban multiplexados con el contenido audiovisual en los sistemas multimedia tradicionales. Sin embargo, en los sistemas multimedia modernos, es posible la generación automatizada y en vivo de los subtítulos, posiblemente por fuentes independientes, así como su personalización (en cuanto a formato, idioma...) y su presentación en entornos multi-pantalla. Es por ello que la sincronización de subtítulos con contenido audiovisual, ya sea en la pantalla principal como en pantallas secundarias, es un problema de investigación actual e interesante.

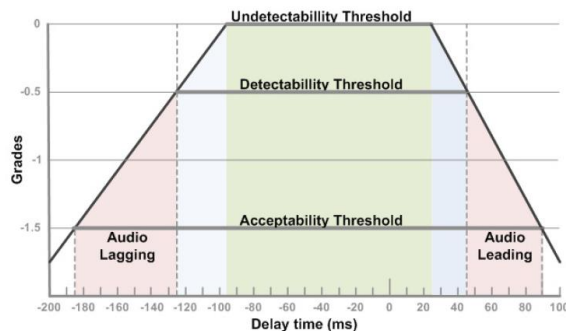


Fig. 4. Umbrales de percepción para lip-sync [19].

Con respecto a la sincronización IDES, varios umbrales de asincronía permisibles se indican en [26]: +/- 10µs para audio altamente acoplado; 2ms para audio en tiempo real; 15ms para audio adelantado y 45ms para video retrasado en *lip-sync*; y +/-80ms para animaciones de video.

Con respecto a la sincronización híbrida, también queda claro que existirán diferentes niveles de asincronía tolerables, en función del caso de uso. Por ejemplo, cuando se presenta contenido web relacionado con el flujo broadcast, los niveles de sincronización necesarios no son muy estrictos, pero si se reproducen videos multi-vista, ya sea en modo mosaico en un mismo dispositivo o en escenarios multi-pantalla, se requiere sincronización a nivel de trama (*frame-accurate synchronization*). Este concepto es relevante, porque va a vincular los niveles temporales de sincronización necesarios a la tasa de envío/reproducción y, por tanto, al ancho de banda consumido por cada componente de video. Por ejemplo, para videos con una tasa de de 25fps, la asincronía no deberá superar los 40ms (1/25 ms). Así pues, se puede deducir que a mayor tasa de tramas, menores serán los niveles de asincronía permisibles. Sin embargo, el concepto de sincronización a nivel de trama es un poco ambiguo, puesto que una asincronía tan pequeña como 1ms ya podría provocar que se estuviesen reproduciendo tramas diferentes de cada flujo durante ese periodo.

Por último, en [20] se recopilan y analizan hasta 20 casos de uso diferentes en los que IDMS es necesaria o ventajosa. Además, dichos casos de uso se clasifican, de manera cualitativa y aproximada, en función de los niveles de sincronización que precisan, según datos recogidos en publicaciones relacionadas. El resumen de esta clasificación se puede consultar en la Tabla II. Se puede apreciar que los umbrales de asincronías permisibles para los sistemas de audio distribuidos son muy estrictos (incluso por debajo del milisegundo). Asimismo, las aplicaciones de multi-conferencia y juegos en red también requieren niveles de sincronización bastante estrictos y, ya en el nivel inferior, aplicaciones como Social TV son más flexibles en este aspecto. Por ejemplo, las evaluaciones subjetivas realizadas en [27] mostraron que 500ms es el nivel inferior de asincronía que empiezan a detectar algunos usuarios en un entorno Social TV distribuido, que asincronías de 1s ya suelen ser perceptibles (más para los usuarios que chatean mediante servicios de audio-conferencia que para los que lo hacen mediante texto) y que asincronías de 2s ya son muy molestas para la gran mayoría de usuarios.

Sin embargo, a pesar de dicha clasificación preliminar, se necesita más investigación con tal de determinar, de manera subjetiva y rigurosa, los límites de asincronías permisibles en diferentes escenarios y aplicaciones IDMS, en función del canal de interacción (texto, audio y video chat) utilizado.

De manera general, se puede apreciar que los niveles de asincronía permisibles para los diferentes tipos de sincronización y tipos de datos multimedia (excepto para aromas) son bastante menores que las magnitudes de diferencias de retardos existentes en las redes actuales. Por tanto, se van a necesitar soluciones precisas y adaptativas para la sincronización multimedia, con el objetivo final de proporcionar una QoE satisfactoria, dados unos niveles de QoS ofrecidos por el entorno de red.

Tabla II
UMBRALES DE ASINCRONÍA PERMISIBLES PARA IDMS

Nivel de sincronización	Requisitos temporales aproximados	Casos de Uso Relevantes
Muy Alto	~ 10 µs – 10 ms	- Altavoces estéreo en red - Array de transductores acústicos en fase - Video walls en red
Alto	~ 10 – 100 ms	- Tele-orquesta distribuida - Juegos de preguntas en red - Juego en red multi-jugador - Conferencia en grupo - Sistemas de conferencia local con amplificación de audio - Aplicaciones multi-pantalla
Medio	~ 100 – 500 ms	- Educación a distancia - Trabajo en grupo en red - Juego de baja interactividad - Votaciones online - Conmutación de la reproducción entre dispositivos
Bajo	~ 500 – 2000 ms	- Compartición de comandos de navegación en TV - TV Social

B. Impacto del ancho de banda

Se ha comentado previamente que algunos tipos y ejemplos de sincronización multimedia requieren niveles de sincronización a nivel de trama. Así pues, la tasa de tramas de los flujos multimedia involucrados y, por tanto, el ancho de banda consumido, influirá en los umbrales de asincronía permisibles.

Asimismo, el envío de datos multimedia con codificación escalable o multi-capa, así como audio multi-canal o espacial, influirá tanto en el ancho de banda necesario como en los niveles de sincronización que se precisan, ya que todos los componentes de las señales multimedia se deberán reproducir de forma sincronizada.

También se intuye que el tamaño y formato de video, el tipo de la pantalla y de dispositivo de consumo pueden influir en la percepción humana. Sin embargo, no existen otros estudios que hayan determinado la influencia de dichos factores sobre los umbrales de asincronía permisibles. Únicamente, en [28] se observa que, en dispositivos móviles, el error de sincronización es más tolerable cuando menor es el tamaño de la imagen (para tasas de tramas reducidas), pero no se proporcionan umbrales.

C. Impacto de la tasa de pérdidas

Es obvio que la pérdida de paquetes tiene un claro impacto sobre la fluidez de la reproducción de los flujos multimedia y, por tanto, sobre los diferentes tipos de sincronización multimedia. Por lo tanto, es interesante determinar el impacto de la pérdida de paquetes y la degradación de video y audio, para diferentes formatos de codificación, sobre la sincronización multimedia y, en definitiva, sobre la percepción de los usuarios (QoE), en diferentes tipos de servicios multimedia.

En este sentido, cabe destacar que en el caso de que los diferentes componentes multimedia se multiplexen en un flujo agregado, las pérdidas de paquetes afectarán a todos los componentes, mientras que esto podría no ocurrir si se envían en flujos separados. Además, las pérdidas de paquetes pueden tener mayor impacto cuando se incluyen datos multimedia discretos (e.g., aromas), porque la pérdida de un

único paquete podría provocar que la MU en cuestión no se pueda presentar al usuario, privando de la inmersividad deseada. Por tanto, en estos casos, el diseño de mecanismos eficientes de recuperación frente a errores es muy deseable.

V. INFLUENCIA DE LOS ESQUEMAS DE CONTROL IDMS

En cuanto a la sincronización IDMS, las soluciones existentes se pueden agrupar según el uso de 3 esquemas de control o arquitecturales diferentes para el intercambio de la información necesaria para conseguir la sincronización (véase la Fig.5) [5, 29]: dos de ellos centralizados (M/S Scheme y Synchronization Maestro Scheme o SMS) y otro distribuido (Distributed Control Scheme o DCS). Por un lado, en [5] se proporciona una comparación cualitativa exhaustiva entre estos 3 esquemas de control considerando varios factores relacionados tanto con la QoS (p.ej., robustez, escalabilidad, sobrecarga de tráfico...) como con la QoE (p.ej., igualdad de condiciones, coherencia...), cuyo resumen se puede consultar en la Tabla III. Por otro lado, esta comparación cualitativa se corrobora mediante una evaluación objetiva en entornos simulados en [29].

VI. INFLUENCIA DE LA SELECCIÓN DE LA REFERENCIA TEMPORAL PARA LA SINCRONIZACIÓN

En los diferentes tipos de sincronización multimedia, cuando se detectan situaciones de asincronía, la primera decisión consiste en seleccionar la referencia temporal a la que sincronizarse. En el caso de sistemas incluyendo audio y video (*lip-sync*), el audio se suele seleccionar como el flujo de referencia, ya que los usuarios son más tolerables a ajustes en video que en audio [17]. Sin embargo, en algunos casos, se selecciona el flujo que experimenta mayor retardo como la referencia, puesto que no es posible reproducir el contenido multimedia que todavía no se ha recibido.

Con respecto a la sincronización híbrida, como los flujos broadband normalmente experimentan mayores retardos que los flujos broadcast, ello implica que el contenido broadcast se debe retrasar hasta recibir el contenido broadband asociado, lo que conllevará problemas de interactividad en servicios de TV. Así pues, es necesario y muy relevante minimizar la latencia en las tecnologías de HTTP streaming.

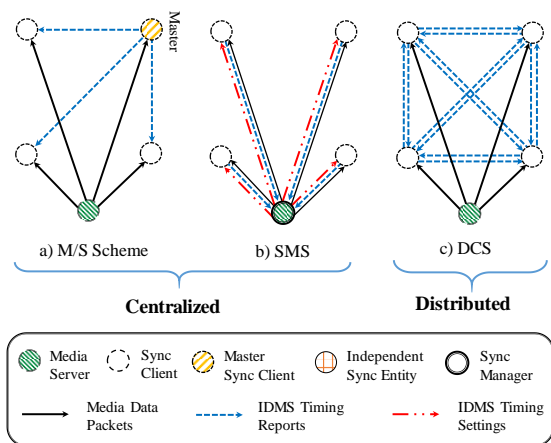


Fig. 5. Esquemas de control IDMS.

Tabla III
COMPARACIÓN ENTRE ESQUEMAS DE CONTROL IDMS

		Factors									
		Robustness	Scalability	Traffic Overhead	Interactivity	Consistency	Causality	Coherence	Flexibility	Fairness	Security
Scheme	M/S	3	1	1	1	2	1	3	3	3	2
	DCS	1	2	3	2	3	3	2	1	1	3
	SMS	2	3	2	3	1	2	1	2	2	1

1: Mejor Esquema; 2: Buen esquema; 3: Peor Esquema

En los sistemas multimedia modernos, que pueden incluir información multimedia multi-sensorial (*multimedia*), como aromas o haptics, de manera adicional a los flujos de audio y video, un tema de investigación muy interesante es determinar las políticas más idóneas para la selección del flujo o el componente multimedia de referencia para la sincronización, tal y como se sugiere en [7].

En el caso de IDMS, si se utilizan los esquemas SMS y DCS, se pueden emplear varias políticas dinámicas para la selección de la referencia temporal para la sincronización [29, 30], con diferentes implicaciones sobre la QoS y la QoE. La selección del receptor más rápido como la referencia puede ser ventajosa para aplicaciones colaborativas, pero puede provocar situaciones de *buffer underflow* si el receptor que marca la referencia es más rápido que ciertos límites permisibles. La selección del receptor más lento como la referencia es apropiada para garantizar igualdad de condiciones (*fairness*) en aplicaciones de ámbito competitivo, pero puede provocar situaciones de *buffer overflow* y retardos excesivos si el receptor que marca la referencia es más lento que ciertos límites permisibles. También se puede seleccionar un punto de reproducción virtual, obtenido como la media de todos los puntos de reproducción involucrados, como la referencia para la sincronización. Esta estrategia permite minimizar el número y la magnitud de los ajustes de sincronización necesarios, pero no puede garantizar que se eviten situaciones de *buffer underflow/overflow*. Otra posibilidad es definir como referencia a un receptor virtual con un punto de reproducción idóneo (sin desviaciones), al que todos los receptores se deberán ajustar. Esta estrategia garantiza, siempre y cuando las condiciones de la red sean estables, una evolución constante del retardo extremo-a-extremo, que es una característica deseable en las comunicaciones multimedia. Finalmente, otras posibilidades son la selección del participante más activo o el que tenga un rol más prioritario (p.ej., el profesor en servicios de *e-learning*) como la referencia para la sincronización, pues seguramente será el foco de atención de la sesión multimedia.

Por último, es importante recalcar que todas estas estrategias de selección de la referencia temporal para la sincronización serán aplicables siempre y cuando tanto los retardos de red como los retardos extremo-a-extremo se mantengan dentro de límites permisibles [6].

VII. CONCLUSIONES

En este artículo se ha descrito la relevancia de los diferentes tipos y ejemplos de sincronización multimedia sobre la QoE, así como su relación con diferentes parámetros de QoS, como son el ancho de banda, la pérdida de paquetes

y, especialmente, los retardos y su variabilidad. Se han proporcionado datos que demuestran que los niveles de asincronías tolerables dependen en gran medida del tipo de sincronización multimedia, del tipo de datos multimedia y del tipo de aplicación bajo estudio. Un conocimiento riguroso y exhaustivo de estos límites es muy importante, pues dará una idea de los niveles de QoS que las redes de distribución de contenidos actuales deberán ofrecer. Asimismo, se ha mostrado que las diferencias de retardos en redes actuales son muy superiores a los niveles de asincronía tolerables, reflejando así la relevancia de la sincronización multimedia. Es por ello que es necesario minimizar, acotar y controlar tanto la magnitud de los retardos como de su variabilidad para proporcionar una QoE satisfactoria. Por último, se ha descrito la influencia de la selección del esquema de control y de la referencia temporal para la sincronización multimedia sobre varios aspectos de QoS y de QoE. La necesidad de más investigación en esta área también se ha identificado y reflejado a lo largo del artículo.

AGRADECIMIENTOS

El trabajo de Mario Montagud se ha llevado durante el disfrute de un contrato postdoctoral financiado por ERCIM 'Alain Bensoussan' Fellowship Programme. El trabajo de la UPV ha sido financiado, en parte, por el FEDER y por el MINECO, bajo el programa de apoyo a la I+D+i en el proyecto con referencia TEC2013-45492-R.

REFERENCIAS

- [1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", IETF Internet Standard, RFC 3550, July 2003.
- [2] Hybrid broadcast broadband TV, <http://www.hbbtv.org/>
- [3] C. Concolato, S. Thomas, R. Bouqueau, J. Le Feuvre, "Synchronized Delivery of Multimedia Content over Uncoordinated Broadcast Broadband Networks", MMSys'12, North Carolina (USA), Feb. 2012.
- [4] R. van Brandenburg, H. Stokking, O. Van Deventer, F. Boronat, M. Montagud, K. Gross, "Inter-destination Media Synchronization using the RTP Control Protocol (RTCP)", RFC 7272, June 2014.
- [5] M. Montagud, F. Boronat, H. Stokking, R. van Brandenburg, "Inter-destination multimedia synchronization: schemes, use cases and standardization", MMSJ, 18(6), 459-482, November 2012.
- [6] M. Montagud, "Design, Development and Evaluation of an Adaptive and Standardized RTCP-based IDMS Solution", PhD Thesis, UPV (Spain), Supervisors: F. Boronat (UPV) and P. Cesar (CWI), March 2015.
- [7] Z. Huang, K. Nahrstedt, R. Steinmetz, "Evolution of temporal multimedia synchronization principles: A historical viewpoint", ACM TOMCCAP, 9, 1s, Article 34, 23 pages, October 2013.
- [8] F. Boronat, R. Mekuria, M. Montagud, P. Cesar, "Distributed Media Synchronization for Shared Video Watching: Issues, Challenges, and Examples", Social Media Retrieval, Springer Computer Communications and Networks, ISBN 978-1-4471-4554-7, 2013.
- [9] J. Jansen, D.C.A. Bulterman, "User-centric video delay measurements", ACM NOSSDAV 2013, pp. 37-42, Oslo (Norway), February 2013.
- [10] ITU-T Rec. G.1050: "Network model for evaluating multimedia transmission performance over Internet Protocol", 2007.
- [11] ITU-T Rec. G. 114: "One-way transmission time", 2003.
- [12] J. Jansen, P. Cesar, D.C.A. Bulterman, T. Stevens, I. Kegel, J. Issing, "Enabling Composition-Based Video-Conferencing for the Home", IEEE Transactions on Multimedia, 13(5), pp. 869-881, October 2011.
- [13] M. O. Van Deventer, H. Stokking, O. A. Niamut, F. A. Walraven, V. B. Klos, "Advanced Interactive Television Service Require Synchronization", IWSSIP 2008, Bratislava (Slovak Rep.), June 2008.
- [14] W. Kooij, H. Stokking, R. van Brandenburg, P-T de Boer, "Playout delay of TV signals: measurement system design, validation and results", ACM TVX 2014, Newcastle (UK), June 2014.
- [15] B. Blesser, "Digitization of audio: A comprehensive examination of theory, implementation, and current practice", J. Audio Eng. Soci., 26(10), pp. 739-771, 1978.
- [16] R. Steinmetz, C. Engler, "Human perception of media synchronization", Technical Report 43.9310, IBM European Networking Center Heidelberg, 1993.
- [17] R. Steinmetz, "Human perception of jitter and media synchronization", IEEE JSAC, 14(1), pp. 61-72, January 1996.
- [18] ITU-R BT.1359, International Telecommunication Union Radio communication Sector Relative Timing of sound and Vision for Broadcasting.
- [19] ATSC Implementation Subcommittee Finding: Relative Timing of Sound and Vision for Broadcast Operations, Doc. ID-191, June 2003.
- [20] A. Mason, R. Salmon, "Factors affecting perception of audio video synchronization in television", BBC R&D White Paper WHP176, 2009
- [21] R. Dannenberg, R. Stern, "Experiments concerning the allowable skew of two audio channels operating in the stereo mode". Personal Communications, 1993.
- [22] L. Goldmann, L. Jong-Seok Lee, T. Ebrahimi, "Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection", IEEE ICIP 2010, Hong Kong, September 2010
- [23] T. Fujimoto, Y. Ishibashi, S. Sugawara, "Influences of Inter-Stream Synchronization Error on Collaborative Work in Haptic and Visual Environments", Symposium on Haptic interfaces for virtual environment and teleoperator systems, March 2008.
- [24] Ghinea, G.; Ademoye, O.A., "Perceived Synchronization of Olfactory Multimedia", Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol.40, no.4, pp.657-663, July 2010
- [25] S. Hoshino, Y. Ishibashi, N. Fukushima, S. Sugawara, "QoE assessment in olfactory and haptic media transmission: Influence of inter-stream synchronization error", IEEE CQR Workshop, May 2011.
- [26] Y. Bang, J. Han, K. Lee, J. Yoon, J. Joung, S. Yang, J. Rhee. "Wireless Network Synchronization for Multichannel Multimedia Services", IEEE ICAT 2009, Sarajevo (Bosnia and Herzegovina), February 2009.
- [27] D. Geerts, I. Vaishnavi, R. Mekuria, O. van Deventer, and P. Cesar, "Are we in sync?: synchronization requirements for watching online video together", ACM CHI 2011, Vancouver (Canada), May 2011.
- [28] I. Curcio, M. Lundan, "Human Perception of Lip Synchronization in Mobile Environment", IEEE WoWMoM, Helsinki, June 2007.
- [29] M. Montagud, F. Boronat, H. Stokking, P. Cesar "Design, Development and Assessment of Control Schemes for IDMS in a Standardized RTCP-based Solution", COMNET, 70(9), pp. 240-259, Sept. 2014.
- [30] M. Montagud, F. Boronat, "Enhanced Adaptive RTCP-based Inter-Destination Multimedia Synchronization Approach for Distributed Applications", COMNET, 56(12), pp. 2912-2933, August 2012.

Estudio de QoS y QoE y Propuestas de Sistemas para la Mejora de los Servicios Multimedia e IPTV

Jose M. Jimenez, Alejandro Canovas, Juan R. Diaz, Jaime Lloret

Inst. de Inv. para la Gestión Integrada de Zonas Costeras, Universitat Politècnica de València
C/ Paranimf, n-1, 46730 Grao de Gandia (Valencia), Spain
jojier@dcom.upv.es, alcalasol@epsg.upv.es, juanramon.diazsantos@gmail.com, jlloret@dcom.upv.es

Resumen- La evolución de las redes y el incremento de la disponibilidad de ancho de banda han permitido incrementar la cantidad de servicios ofertados por los operadores de telecomunicaciones así como la calidad de estos. La rápida implementación de servicios multimedia y redes IPTV ha llevado a la necesidad de realizar múltiples estudios sobre los parámetros de calidad de servicio (QoS) y estimaciones de calidad de experiencia (QoE) para mejorar dichos servicios. El presente trabajo muestra los diferentes estudios y propuestas de sistemas para la mejora de distribución de flujos Multimedia e IPTV realizadas por el grupo de investigación Comunicaciones y Redes del Instituto de Investigación para la Gestión Integrada de Zonas Costeras de la Universidad Politécnica de Valencia. Se presentan los mayores logros y conclusiones obtenidas durante este proceso, así como las líneas de investigación actuales.

Palabras Clave- IPTV, QoS, QoE, redes multimedia

I. INTRODUCCIÓN

Desde hace unos años, de forma estratégica, los operadores y proveedores de servicios están adaptando sus redes para proveer servicios con una calidad de servicio (QoS) adecuada. QoS se puede definir como la capacidad de la red para ofrecer un servicio mejor o especial a un conjunto de usuarios o de aplicaciones o a ambos, en detrimento de otros usuarios o aplicaciones, o ambas cosas. Las redes multimedia tienen una serie de requerimientos especiales y es muy importante garantizar una buena calidad del servicio que perciban los usuarios, es decir la calidad de experiencia del usuario (QoE). La UIT define QoE [1] como la aceptabilidad global de una aplicación o servicio tal y como se percibe subjetivamente por el usuario final. Incluye la totalidad de efectos del sistema extremo a extremo (cliente, terminal, red, servicios de infraestructura...) y puede verse influenciada por las expectativas de los usuarios y el contexto. Esto tiene como consecuencia que la QoE se mida subjetivamente y pueda diferir de un usuario a otro. Además, la señal de video digital necesitaría un gran ancho de banda si se quisiera transmitir en su formato original. Por lo tanto, resulta necesario realizar una compresión previa a la transmisión para lograr una eficiencia en el almacenamiento y para distribuirla a través de las redes IP.

IPTV es un servicio ofertado por los operadores que hace uso de conexiones de banda ancha y redes de próxima generación como plataforma para entregar servicios Triple Play [2] (servicios de datos, voz y vídeo). Se considera vital la calidad del servicio de IPTV que percibirán los usuarios.

Existen informes de empresas del sector que confirman las tendencias de los usuarios para los próximos años. En el Cisco Visual Networking Index (VNI) 2012-2017 [3] se

prevé que el tráfico IP global se multiplicará por tres entre 2012 y 2017 y que los usuarios generarán 3 billones de minutos de vídeo al mes en Internet a escala global. Además, en el informe Cisco VNI Service Adoption Forecast 2013-2018 [4] se predice que a nivel mundial la televisión digital crecerá de 1000 millones de hogares en 2013 hasta 1500 millones de hogares en 2018, y que el vídeo bajo demanda (VoD) crecerá de 306 millones de suscriptores en 2013 a 451 millones de suscriptores en 2018.

Debido a todas las premisas que plantemos, resulta obvia la necesidad de realizar estudios en ámbitos como la codificación, tecnologías de transmisión, QoS y QoE, para poder llevar a cabo una entrega de servicio de IPTV efectiva.

El presente trabajo muestra los estudios efectuados por el grupo de investigación comunicaciones y redes, del Instituto de Investigación para la Gestión Integrada de Zonas Costeras (IGIC) de la Universidad Politécnica de Valencia a lo largo de los años, que han sido publicados a través de las actas de diferentes congresos o en revistas. En ellos se han realizado investigaciones y pruebas con códecs de video, cuantización, transcoding, color, y también medidas con redes de diferentes tecnologías. Además, mostramos nuestras propuestas para la mejora de QoS y QoE, y las nuevas investigaciones realizadas en el campo de las redes multimedia.

El resto del trabajo se organiza de la siguiente manera. En la Sección II se presentan los trabajos relacionados con medidas respecto a códecs, transcoding, cuantizador, color, video sobre redes inalámbricas, etc. En la Sección III se presentan los estudios y propuestas de sistemas para la mejora de la QoS y QoE.

II. INVESTIGACIONES PREVIAS Y PRUEBAS INICIALES

Desde el año 2005 el grupo ha realizado diversas investigaciones y pruebas iniciales, obteniendo medidas respecto a los códecs, transcoding, cuantizador, y color, tanto en redes cableadas como inalámbricas, que van a ser presentadas a lo largo de esta sección.

En [5] realizamos una comparativa de los resultados obtenidos al comprimir vídeos con los códecs DivX, XviD y H.264 según la variación de la luminancia (capturados a diferentes horas del día) para su aplicación al campo de la vídeo vigilancia. Se compararon en función del factor de compresión, del tiempo necesario para comprimir un seg. de vídeo, del porcentaje de tiempo de procesador y el número de entradas y salidas (E/S) que se tienen en el procesador al realizar la compresión. Se tuvieron en cuenta diferentes tasas de bits para cada uno de los códecs. Se demostró que la elección del códec dependía de las limitaciones o

características del sistema implementado. Si las limitaciones del sistema estaban relacionadas con el ancho de banda de la red, el mejor resultado era el presentado por el códec XviD. Si se necesitaban tiempos de compresión cortos, los mejores eran los códecs DivX y XviD. En caso de que el propio dispositivo de captura fuera la limitación del sistema, porque tenía más procesos en funcionamiento o porque el dispositivo debía capturar desde varias cámaras, la mejor elección era el códec H.264. Este proceso se muestra en la Fig. 1.

En nuestro siguiente trabajo [6] mostramos los parámetros que debíamos tener en cuenta y las medidas adoptadas para desarrollar redes inalámbricas orientadas a vídeo vigilancia en entornos rurales. Inicialmente se estudiaron diferentes códecs XviD, DivX y H.264. Entonces, se tomaron decisiones respecto a la transmisión teniendo en cuenta las limitaciones del sistema implementado o sus características. En este entorno se obtuvieron resultados similares a los que presentamos previamente en [5]. Se propuso un diseño de red inalámbrica basado en el estándar IEEE 802.11g. Además, destacamos que la implementación de este tipo de redes podía servir para otros entornos como la observación de animales, detección de intrusos, procesamiento en tiempo real para la detección y verificación de incendios en los bosques [7], controlar el impacto de los visitantes en los parques naturales y ofrecerles caminos alternativos para visitar, etc.

En [8], publicado en 2006, se explicó cómo crear un video con propiedades de alta definición sin utilizar cámaras de vídeo HDTV ni dispositivos de televisión de alta definición. Después se discute sobre el códec adecuado para la compresión de HDTV, manteniendo sus características. Se midió la ganancia de compresión, el tamaño del archivo comprimido y el tiempo que se necesitaba para comprimir un segundo de vídeo. Posteriormente se publicó una extensión del trabajo anterior en [9]. Profundizando en el estudio se detectaron algunas anomalías para el códec XviD en HDTV (el rendimiento del códec es mayor a 2000 Kbps), ya que comprime más el archivo que el valor esperado.

En [10] propusimos un sistema que elegía el mejor tipo de red inalámbrica de acceso para los clientes IPTV. Los dispositivos podían ser de doble banda o tri-banda. La elección se realizaba en función de los requerimientos, las redes disponibles y algunos parámetros de red como el número de paquetes perdidos o el retardo, ofreciendo la máxima calidad de experiencia al cliente. Se tomaron medidas en diferentes tipos de redes inalámbricas.

En [11] realizamos estudios sobre el factor de escala para la cuantificación (QS). Analizamos el QS como un parámetro de calidad visual. En este trabajo calculamos QS en el decodificador, extrayendo la información codificada en los paquetes MPEG-2 TS, con la finalidad de crear un modelo de referencia empleado en el cálculo de la QoE. El artículo muestra medidas de calidad de vídeo y del ancho de banda consumido por vídeos con diferentes calidades. La Fig. 2 muestra la calidad de vídeo en IPTV usando el cuantificador propuesto. Demostramos que un factor QS óptimo puede emplearse para ahorrar ancho de banda en una red IPTV o para mejorar la calidad de vídeo consumiendo el mismo ancho de banda.

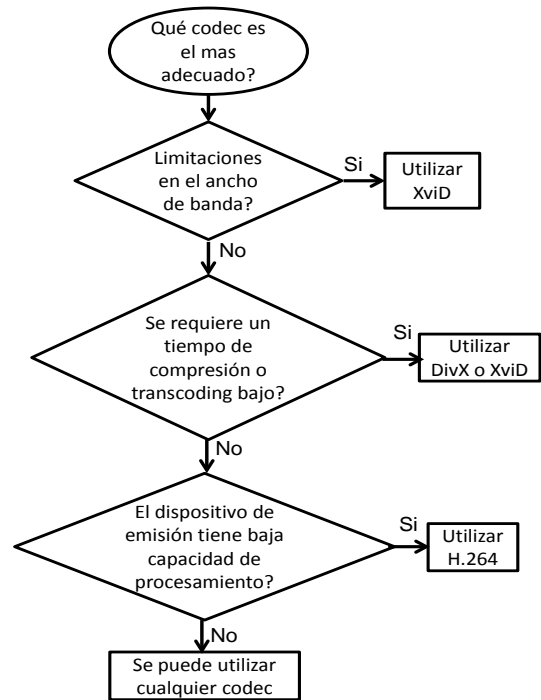


Fig. 1. Proceso de elección del códec según las características del sistema

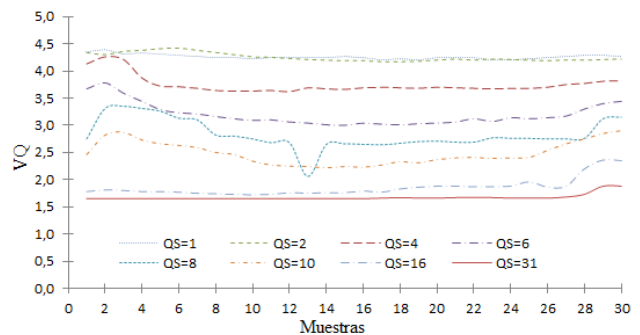


Fig. 2. Calidad de vídeo en IPTV usando el cuantificador propuesto

Posteriormente en [12] extendemos este trabajo llegando a la conclusión de que QS4 y QS6 son los valores óptimos cuando incluimos los cambios en la red y las condiciones de la red no son óptimas. Además, en este trabajo se han añadido tests de ancho de banda, jitter, retardo y pérdida de paquetes con el fin de probar el funcionamiento del canal IPTV y la VQM (Video Quality Metric) y el MOS (Mean Option Score).

En [13] estudiamos la transmisión de vídeo 3D cuando las 2 imágenes son enviadas dentro del mismo canal. Probamos que podía funcionar realmente en las redes IP existentes. Testeamos el ancho de banda requerido para la transmisión del vídeo con diferentes tasas de bits y los resultados fueron satisfactorios. También testeamos diferentes niveles de codificación para una adecuada calidad de vídeo 3D, obteniendo buenos resultados. Observamos valores muy bajos en los parámetros de jitter y retardo (véase la Fig. 3), lo cual asegura una QoS adecuada. Los peores resultados se obtuvieron en la pérdida de paquetes, pues mostraron un alto impacto en la calidad del vídeo recibido cuando variamos del 0,1% a 2%, y del 2% a 5%. Por este motivo consideramos que un punto importante debería ser mejorar el sistema de transcodificación, con el fin de disminuir el efecto de la pérdida de paquetes.

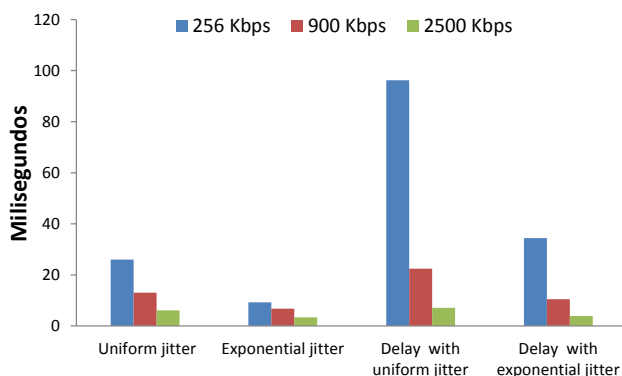


Fig. 3. Valores de jitter y retardo para varios bitrates

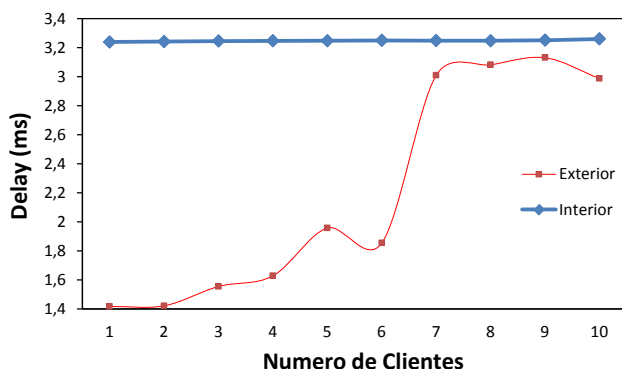


Fig. 4. Retardo medio para 1 canal multicast



Fig. 5. Interfaz gráfica

En [14] probamos el rendimiento de la IPTV cuando se utiliza IEEE 802.11n y proporcionamos los parámetros de QoS medidos en 3 bancos de prueba con hasta 17 clientes de IPTV. Nadie había testeado el rendimiento de IPTV a través de IEEE 802.11n hasta ese momento y el trabajo podía ser utilizado por los proveedores de servicios de IPTV a fin de conocer el comportamiento de la IPTV en una red de acceso inalámbrico. Probamos en dos ambientes principales: interior y exterior. Observamos que el retardo, jitter y pérdida de paquetes no aumentan linealmente con el número de usuarios, tienen un comportamiento diferente en cada caso. La Fig. 4 muestra el retardo medio para 1 canal multicasta en ambos casos. En las pruebas de interior, la red IPTV gestionaba una docena de clientes sin afectar al ancho de banda total; mientras que en las pruebas al exterior, se observó que con 12 clientes aumentaba significativamente la media del ancho de banda.

En [15] nos centramos en el diseño, la implementación, análisis y evaluación de un algoritmo de gestión de red, que monitoriza la red para asegurar la calidad de los servicios de video 3D. Para llevar a cabo el test de rendimiento del algoritmo, realizamos diferentes pruebas con el fin de obtener valores de referencia de retardo, jitter, pérdida de

paquetes y MOS. Éstos se utilizaron para analizar el sistema de codificación en términos de calidad de servicio para 3D/HD IPTV. Las técnicas de codificación 3D utilizadas en fueron codificación de vídeo estéreo y codificación 2D+Z. Se propuso un algoritmo para hacer una transmisión efectiva de vídeo estereoscópico que se basaba en el estado de la red y el uso de vídeos codificados varios con varios ratios de bits y codecs, entregando el mejor vídeo en cada momento.

En [16] presentamos una nueva herramienta para realizar pruebas de rendimiento de la red. Se ha desarrollado una aplicación que una vez ejecutada, se puede elegir que sea cliente o servidor. El cliente permite elegir el tipo de tráfico que envía al servidor (Audio o Video) y el servidor reenvía dicho tráfico hasta el Cliente. En el cliente se calcula y registra el jitter, delay, mensajes perdidos, porcentajes de mensajes perdidos, ancho de banda, etc. La aplicación muestra gráficamente todos estos datos por segundo en una interfaz que se maneja de forma intuitiva. No es capaz de filtrar las medidas por flujos, pues muestra sólo los valores instantáneos de la comunicación configurada entre el servidor y el cliente. La Fig. 5 muestra la interfaz gráfica de la herramienta. Las medidas de jitter, delay, paquetes perdidos y ancho de banda se calculan sin la necesidad de que exista ninguna cabecera RTP (la mayoría de programas, como por ejemplo Wireshark, requieren de RTP para poder calcularlo, en caso contrario, no son capaces de hacerlo). Para poder probar su efectividad realizamos una serie de medidas sobre una red real. Comparándola con las herramientas existentes, podemos decir que la mayor contribución está en la creación de la primera herramienta gráfica para medir el rendimiento de red que ofrece resultados y permite almacenarlos para un posterior estudio.

En [17] proponemos un esquema de selección de red inteligente (INS) para tomar decisiones de handover vertical en VANET, cuando se envían flujos de video o audio, utilizando comunicaciones V2I. El esquema propuesto fue diseñado basado en la maximización de la función de puntuación para clasificar los candidatos disponibles de red inalámbrica. Se utilizaron tres parámetros de entrada para desarrollar la función de puntuación de maximización, que fueron: pérdidas en la relación señal-ruido, capacidad residual de canal y tiempo de vida de conexión. Los resultados de las simulaciones muestran que el esquema INS propuesto mejora los enfoques existentes en cuanto a la disminución de la probabilidad de traspasos innecesarios, averías de conexión de enlace, y los fracasos de traspaso. El esquema también disminuyó el retardo medio de handover, el retardo de los paquetes de extremo a extremo para VoIP y aplicaciones vídeo, el índice de pérdida de paquetes de VoIP y aplicaciones vídeo, además de hacer el proceso de selección de red más eficiente.

En [18] proponemos un esquema de caracterización de canal combinado con un control de admisión dinámico en una red de sensores multimedia, para compartir los recursos de red entre nodos. Analizamos y mostramos el comportamiento de los nodos que utilizan diferentes tecnologías de acceso a la red y los efectos de canal para cada tecnología. También se muestra la existencia de tasas óptimas de la llegada nodos con el fin de mejorar el uso de control de admisión dinámica cuando se utilizan los recursos de red. Presentamos múltiples simulaciones para validar la eficacia del control de admisión dinámico propuesto.

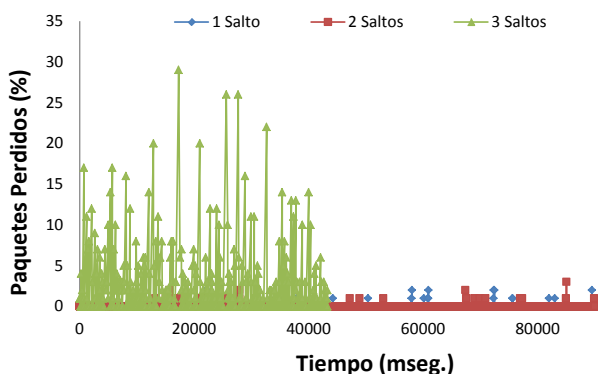


Fig. 6. Pérdida de paquetes, vídeo a 1800 kbps con 1, 2 y 3 saltos.

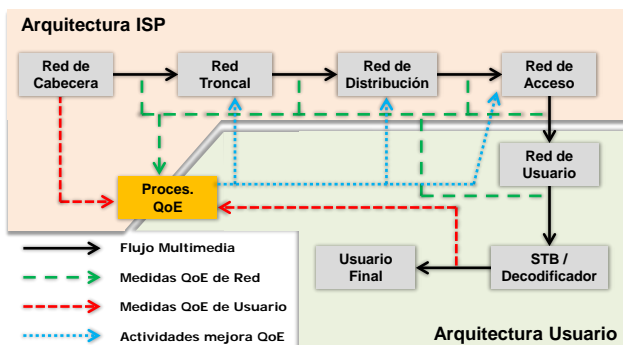


Fig. 7. Sistema de gestión de red basado en el QoE

TABLA 1. VALORES OBTENIDOS DEL ANCHO DE BANDA

Ancho de banda	Sin Transcoding	Con Transcoding	Diferencia (Reducción)
Mínima	3.3 Mbps	1.63 Mbps	50.6%
Máxima	8.5 Mbps	5.35 Mbps	37%
Media	6.2 Mbps	5.2 Mbps	16.1%
Varianza	2.92 Mbps	0.38 Mbps	86.9%

En [19] proponemos un nuevo protocolo que permite el intercambio de servicios multimedia en la red ad hoc y tiene en cuenta los parámetros de QoS y QoE para seleccionar el mejor nodo proveedor de servicios multimedia. Mostramos el protocolo diseñado y los algoritmos de decisión. Por un lado, el sistema tiene en cuenta la latencia, jitter, paquetes perdidos y los parámetros de ancho de banda con el fin de seleccionar el mejor nodo proveedor de servicios. Por otro lado, el sistema tiene en cuenta un parámetro estimado de QoE y el nodo más cercano, que implica menos RTT y que los tiempos de zapping sean más bajos, con el fin de tener la mejor QoE. Medimos los parámetros de QoS para VoIP, utilizando los códecs de audio G.729 y G.711, para 1, 2, 3 o 4 saltos, y para streaming de video en diferentes ratios de bits. Los valores obtenidos en la prueba muestran que es posible entregar los flujos de vídeo con baja resolución. Para la transmisión de vídeo en el rango de 900 a 1800 Kbps, los parámetros de Q muestran que no es posible entregar el video cuando existen 3 o 4 saltos con un nivel de calidad aceptable. La Fig. 6 muestra la pérdida de paquetes en la transmisión de un vídeo a 1800 Kbps para 1, 2 y 3 saltos. Hay valores superiores a varios segundos para la latencia y jitter.

En [20] presentamos el estudio de diferentes codificaciones de un vídeo en función de la tasa de bits y su tiempo de codificación. El estudio nos permite conocer las características de los formatos de vídeo MPEG-4, Ogg y WEBM. En términos de tamaño del archivo comprimido, los valores obtenidos para PAL son ligeramente mayores que

para el sistema NTSC. Esto es debido a la diferencia de la velocidad de fotogramas entre los sistemas. Por lo tanto, los valores de tamaño son algo más altos que en el sistema PAL tanto en MP4 (H.264) como en Ogg (Theora), mientras que en WebM (VP8) son prácticamente idénticos en PAL y NTSC. Consideramos que la opción más favorable, si se tiene en cuenta la calidad de vídeo, el tamaño del archivo y el tiempo de compresión, es el formato WebM, debido a que su tamaño final es el más bajo, con un tiempo de codificación más bajo y con una alta calidad próxima a la de los bitrates más altos de MP4.

III. SISTEMAS PARA LA MEJORA DE QoS Y QoE

En esta sección presentamos las propuestas de sistemas para la mejora de la QoS y QoE que hemos desarrollado.

En [21] propusimos un sistema de gestión para IPTV en función de la QoE del usuario. Dicho sistema garantizaba la mejor QoE independientemente del tipo de conexión, cableada o inalámbrica. Comparamos las medidas de retardo, jitter, ancho de banda y tiempo de zapping en redes IEEE 802.11 b/g y IEEE 802.3u, para conocer las diferencias entre los parámetros de QoE en ambas redes. Observamos que aunque la red IEEE 802.3u era más estable, existía muy poca diferencia entre ambas, solamente se apreciaba una diferencia significativa en el jitter. Propusimos una fórmula basada en los parámetros de QoE de los usuarios, de esta forma se permitía realizar en la red todas las operaciones adecuadamente. Nuestro sistema calculaba los parámetros de QoE, tanto desde el punto de vista de la red como del usuario, en cualquier instante y toma decisiones respecto a la tarea apropiada a realizar. La Fig. 7 muestra el sistema de gestión de QoE propuesto para servicios IPTV.

En [22] estudiamos la transcoding para controlar la congestión en las redes de IPTV. En dicho artículo propusimos un sistema que se adaptaba las variaciones de tráfico de un servicio de IPTV, con el fin de evitar la congestión de la red y las pérdidas de paquetes, que afectan a la calidad visual del usuario y por tanto a la QoE. Implementamos un transcoding de vídeo como técnica alternativa para evitar la saturación de tráfico, con el fin de reducir la tasa de bits de un flujo de vídeo, la tabla 1 muestra valores estadísticos del ancho de banda ahorrado.

Posteriormente, en [23] proponemos un algoritmo de gestión de red, basado en medidas de red obtenidas de múltiples pruebas realizadas con anterioridad sobre QoE en IPTV. El algoritmo propuesto se adapta las variaciones de tráfico de un servicio IPTV e incorpora un transcoding de vídeo para reducir la tasa de bits de un flujo de vídeo cuando es necesario. En nuestras pruebas hemos comprobado que el sistema propuesto se consigue una adaptación de vídeo muy eficaz, atenuando las ráfagas de vídeos y consiguiendo un menor consumo de ancho de banda en la red. Aunque, esta reducción produce una calidad visual inferior, cuando la reducción en la tasa de bit no supere un cierto valor, no es perceptible por el usuario.

A partir del trabajo realizado en [13], en [24] propusimos un algoritmo para tomar las decisiones adecuadas en una red IPTV cuando se transmite vídeo estereoscópico. Tomamos medidas de ancho de banda, retardo y jitter con el fin de evaluar la QoS y la QoE del servicio ITPV y demostrar la necesidad utilizar códecs de compresión estereoscópico de alto rendimiento.

TABLA 2. VALORES MEDIOS DE

	Average uniform jitter (ms)	Average uniform delay (ms)	Average exponential jitter (ms)	Average exponential delay (ms)
DivX	3.804	2.428	1.849	2.454
MPEG4. mkv	3.825	2.453	4.652	2.789
MPEG4/H.264	3.845	2.465	1.702	1.860
Xvid	3.806	2.446	1.885	2.463

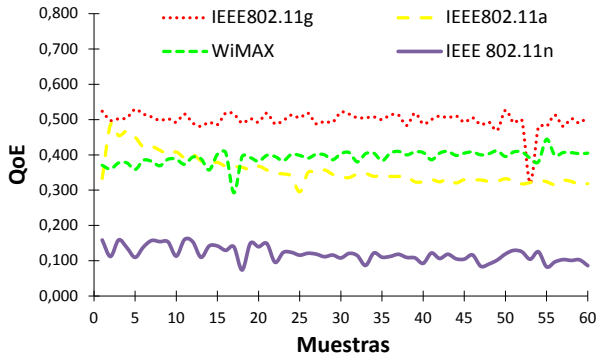


Fig. 8. QoE observada en cada tecnología inalámbrica

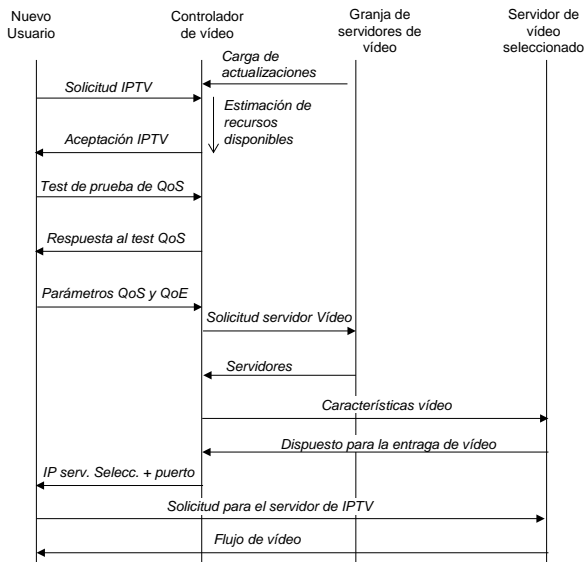


Fig. 9. Protocolo propuesto

No se observó pérdida de paquetes intrínseca y ni desincronización cuando se transmitieron los videos estereoscópicos, pero, los requisitos de ancho de banda de red eran mayores en la red de IPTV debido a la transmisión simultánea de flujos de video estereoscópicos. Para mejorarlo, propusimos la creación de un algoritmo que gestionara la red. Para lograr este objetivo, se midieron los rendimientos de DivX, XviD, MPEG-4 parte 2 y MPEG-4 / H.264 parte 10 para diferentes ratios de bits con el fin de ver cuál funciona mejor en cada situación. También se tomaron medidas de tiempo para codificar un segundo de video, el factor de codificación, el número de operaciones de datos de E/S, el % de tiempo de procesamiento y el número de bytes de datos de E/S. Estos datos empíricos, unidos con algunas consideraciones analíticas, se utilizaron para crear el algoritmo para transmisión de video estereoscópico. El análisis de parámetros jitter y retardo muestra su buen desempeño, pues tienen valores muy bajos, lo que asegura una adecuada QoS. La tabla 2 muestra los valores de jitter y delay cuando en la red se introduce jitter y delay, uniforme y exponencial respectivamente, cuando se está transmitiendo con un ratio de bits de 3129 Kbps.

En [25] propusimos un algoritmo y un protocolo que tenían en cuenta los parámetros de QoS en una red IPTV. Se garantizaba en cualquier instante los valores más altos de QoS para el usuario. Las medidas reales demostraron que el handover vertical realizado por el usuario se producía sin que el propio usuario se percibiera del cambio. Aunque había más paquetes perdidos, y un mayor retardo y jitter, la itinerancia se produjo sin que implicara un alto impacto en la calidad del video recibido por el usuario final (la degradación de la calidad video no fue apreciada por el usuario). Observamos que los casos más graves ocurrían cuando estaba involucrada la tecnología IEEE 802.11a. Posteriormente en [26] proponemos un nuevo algoritmo de red para la distribución de IPTV utilizando tecnologías WiMAX y WLAN. El algoritmo selecciona la tecnología de acceso inalámbrico apropiada con el fin de obtener la mayor QoE para el cliente IPTV. También se explica el protocolo propuesto para su funcionamiento. Mediante las mediciones obtenidas comparamos el retardo, jitter y pérdida de paquetes recibidos por el dispositivo del cliente cuando el sistema selecciona cada una de las tecnologías incluidas en nuestro sistema. También mostramos el ancho de banda efectivo y el ancho de banda consumido por un canal de HDTV en IPTV. En términos de ancho de banda, lo más apropiado es utilizar WiMAX, pero IEEE 802.11g proporciona mejores resultados de QoE. La tecnología IEEE 802.11n presentó los peores resultados. Cuando probamos el impacto, habiendo roaming entre diferentes tecnologías, en el usuario cuando el dispositivo está en itinerancia, hemos observado que el mayor retardo se produce cuando la tecnología IEEE 802.11a está involucrada en la itinerancia, pero en ningún caso era apreciable por el usuario final mientras veía la televisión. La Fig. 8 muestra la QoE observada en cada caso. El parámetro QoE de red, es decir, cómo se ve afectado el QoE al variar los parámetros de red, se definió de manera empírica tras la realización de múltiples experimentos para cada caso específico. En este caso realizamos los test utilizando la fórmula 1.

$$QoE \approx \frac{1}{(Delay + K_1 \cdot Jitter) \cdot e^{PacketLoss}} \quad (1)$$

En [27] presentamos un protocolo y un algoritmo de gestión de red para redes ad hoc basadas en la QoE del usuario final. La Fig. 9 muestra el protocolo propuesto. El algoritmo de gestión de red utiliza la información recibida por el usuario, con el fin de seleccionar el mejor servidor de video (de una granja de servidores) para transmitir contenido IPTV, y seleccionar las características de video que proporcionan la más alta QoE al cliente.

En [28] proponemos una nueva arquitectura, para organizar y gestionar redes ad hoc, basada en clusters para transmitir flujos multimedia. Se trata de una arquitectura que utiliza perfiles de QoS para optimizar el tráfico multimedia. La arquitectura proporciona una solución flexible con la capacidad de garantizar la calidad de la comunicación multimedia a través de la red inalámbrica ad hoc. Es capaz de adaptarse a muchas configuraciones de red física a través de la selección adecuada de los Multimedia Init Profiles (MIP). Se muestra cómo los parámetros de QoS y las características del códec multimedia afectan a la topología del cluster. Por otra parte, el diámetro del cluster afecta gravemente al retardo y jitter. La arquitectura propuesta proporciona un

mecanismo de control para construir la topología apropiada para cada grupo. Además, el sistema utiliza un esquema de reserva de recursos para garantizar la calidad de los flujos multimedia. La Fig. 10 muestra la máquina de estado finito para la arquitectura propuesta.

En [29] proponemos y desarrollamos un nuevo protocolo de comunicación que crea grupos de nodos ad hoc basados en las características del flujo multimedia que se entregan dentro de una red inalámbrica de sensores. Con el fin de lograr este objetivo, hemos definido el MIP como un esquema lógico que nos permite gestionar los requisitos de calidad de servicio y las características de los nodos sensores que construyen el clúster. El protocolo permite la creación de clusters con un diámetro máximo, que es adecuado para cada tipo de flujo multimedia y selecciona los nodos más apropiados, con suficientes recursos, para estar en el camino de la entrega multimedia. La Fig. 11 muestra la comunicación que se puede establecer entre los diferentes elementos que forman un cluster. En el artículo detallamos las características de protocolo, los mensajes diseñados, y las variables utilizadas. Por otra parte, explicamos los procesos de la arquitectura, que detallan cómo se tiene lugar el descubrimiento de vecinos, la creación de vecindad, y la entrega multimedia. Finalmente hemos medido varios casos en un banco de pruebas con dispositivos reales. Hemos demostrado que el protocolo es capaz de alcanzar los valores adecuados de los parámetros de QoS para diferentes MIPs.

En [30] proponemos un nuevo mecanismo de tolerancia a fallos, utilizando como base la arquitectura MWAHCA (presentada en [28]), con el fin de recuperar los flujos multimedia cuando cruzan la red inalámbrica ad hoc y falla un nodo. En este nuevo artículo incluimos un nuevo algoritmo de tolerancia a fallos. Este mecanismo utiliza una ruta de conmutación rápida temporal para mejorar los parámetros de QoS y QoE en la transición de recuperación. En este proceso los paquetes multimedia se reenvían a través de una nueva ruta, que puede no ser la ruta óptima y puede que algunos nodos en el camino no tengan hecha correctamente una reserva de recursos. La Fig. 12 muestra el tráfico multimedia atravesando una red ad hoc utilizando el camino óptimo y cuando se activa Fast Switching Paths (FSP) después de haber fallado un nodo. Con el nuevo mecanismo de tolerancia a fallos, el tiempo de convergencia se reduce significativamente y la cantidad de paquetes perdidos se minimiza. Mientras que los paquetes multimedia siguen siendo remitidos a través de la ruta de conmutación rápida, el mecanismo de tolerancia a fallos construye un tercer camino óptimo, con el fin de mejorar otros parámetros como latencia y jitter. La última ruta se construye siguiendo el proceso de reenvío original proporcionado por la arquitectura MWAHCA. Por lo tanto la reserva de recursos se realiza en todos los nodos en el camino. Los resultados en el laboratorio muestran que el algoritmo tolerante a fallos funciona como se esperaba, pues permite a un flujo multimedia cruzar la red ad hoc con un tiempo de convergencia mínimo después del fallo de un nodo. Los parámetros de calidad de servicio medidos indican que ambos servicios multimedia unidireccionales y bidireccionales no son interrumpidos en ningún momento.

En [31] presentamos un algoritmo de decisión que selecciona el códec más eficiente para comprimir un vídeo en función de sus características de croma, que posteriormente se transmitirá a través de la red IP.

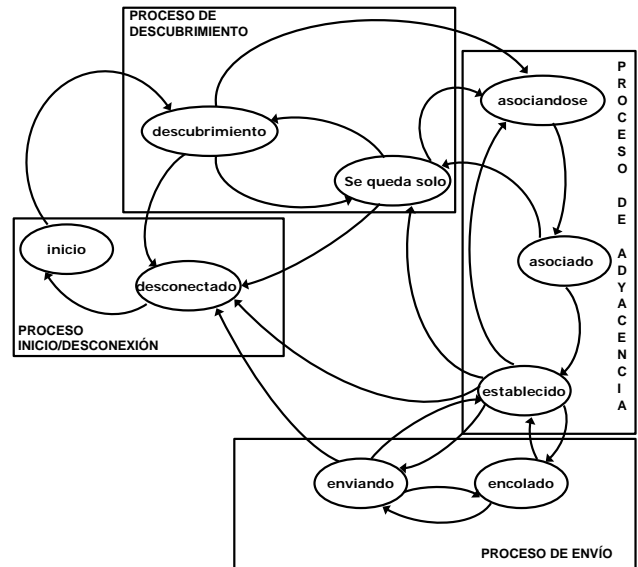


Fig. 10. Máquina de estado finito de la arquitectura

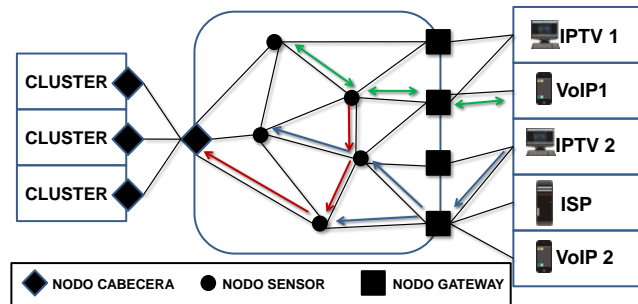


Fig. 11. Elementos del cluster y posibles comunicaciones.

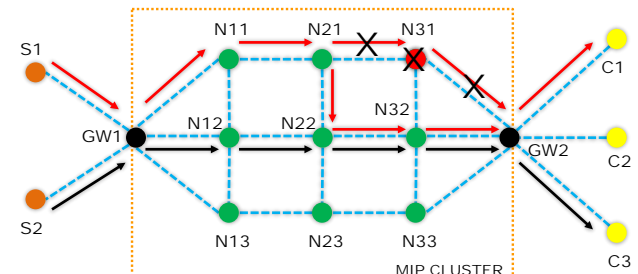


Fig. 12. 2 flujos multimedia atravesando la red ad hoc por el camino óptimo y se activa FSP después de haber fallado un nodo

Teniendo en cuenta las consideraciones de QoS que proporciona la red y la QoE de los usuarios finales, el algoritmo toma la información sobre el espectro de colores del vídeo para seleccionar el códec que se debe utilizar para ahorrar ancho de banda. Realizamos dos pruebas, la primera analizaba vídeos preparados con fondo liso de diferentes colores (negro, blanco, rojo, azul y verde), la segunda prueba ha analizado vídeos reales con colores predominantes con el fin de comprobar la calidad de imagen y el tiempo de compresión. Se observó que diferentes colores sufren diferentes niveles de compresión, y el tiempo en este proceso también depende del color dominante del vídeo. La adaptación dinámica del códec utilizado en la transmisión de vídeo en función de los cambios de color implica un significativo ahorro de ancho de banda de la red y la mejora de la QoE percibida por los usuarios finales. También se propuso un protocolo de control para sincronizar el emisor y el receptor cuando va a tener lugar la entrega de vídeo.

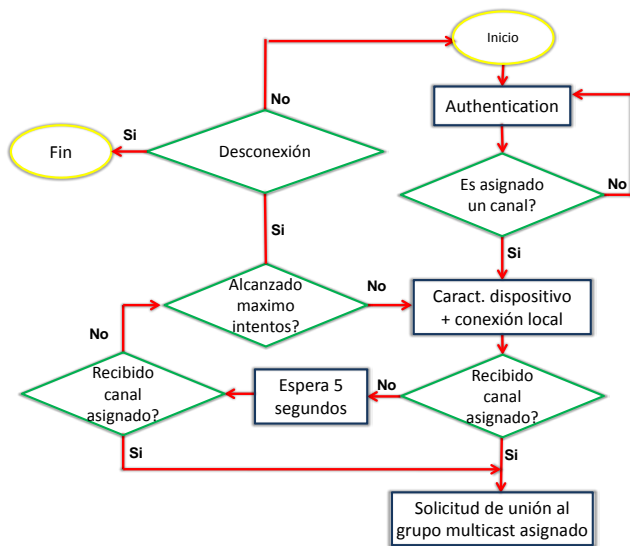


Fig. 13. Algoritmo propuesto para mejorar la QoE en IPTV

En [32] presentamos la propuesta de un nuevo protocolo y algoritmo para mejorar la QoE de los clientes de servicios de IPTV. La Fig. 13 muestra el algoritmo propuesto para mejorar la QoE en IPTV. Inicialmente se llevó a cabo un estudio utilizando una topología básica sobre la que se transmite un vídeo, en la que variamos algunos parámetros y las condiciones de la red. La calidad de vídeo recibido se evaluó utilizando MOS. Después de realizar las pruebas, y estudiar los resultados, podemos afirmar que el principal problema para un operador que presta servicio de IPTV se produce cuando no puede tener el ancho de banda requerido para todos sus clientes. En el artículo proponemos un nuevo algoritmo donde suponemos que los operadores pueden controlar correctamente el retardo, jitter, y ancho de banda para la transmisión de vídeo en función del dispositivo en el que el cliente va a recibir la transmisión. Un paso clave para asegurar que la QoE del usuario final es satisfactoria, es entregar el vídeo con la calidad suficiente de acuerdo a las características del dispositivo. Enviar el vídeo con mayor calidad es innecesario, ya que se desaprovecha el ancho de banda, mientras que si transmitimos a menor calidad de la esperada, disminuirá el QoE en dispositivo del cliente. Por este motivo proponemos tener los vídeos con diferentes codificaciones (o transcódicar) de acuerdo con los dispositivos utilizados por los clientes, y que cuando estos accedan a la red, lo hagan a diferentes canales en función de sus características.

REFERENCIAS

[1] Definition of Quality of Experience (QoE). Pagina Web UTL. <http://www.itu.int/md/T05-FG.IPTV-IL-0050/es>

[2] C. Hellberg, D. Greene, T. Boyes, "Broadband network architectures: designing and deploying triple-play services", Prince Hall PTR Upper Saaler River, NJ, USA, 2007.

[3] Cisco Visual Networking Index. Web en: <http://www.cisco.com/web/ES/about/press/2013/2013-05-29-mitad-poblacion-mundial-conectada-a-internet-en-2017.html> [Consulta 04-05-2015].

[4] Cisco VNI Service Adoption Forecast. Web en: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/vni-service-adoption-forecast/Cisco_VNI_SA_Forecast_WP.pdf [Consulta 04-05-2015].

[5] R. Herrero Doñate, J. Lloret Mauri, J. M. Jiménez Herranz, Estudio de Códecs de Compresión MPEG-4 para su Aplicación a la Videovigilancia, XX Symposium Nacional de la Unión Científica

Internacional de Radio (URSI2005), Gandia (Valencia), 14-16 de Septiembre de 2005.

[6] J. Lloret, P. V. Mauri, J. M. Jimenez, J. R. Diaz, 802.11g WLANs Design for Rural Environments Video-surveillance, International Conference on Digital Telecommunications (ICDT'06), Cap Esterel, Costa Azul (Francia), 26-31 de Agosto de 2006.

[7] J. Lloret, M. Garcia, D. Bri, S. Sendra, A wireless sensor network deployment for rural and forest fire detection and verification, sensors 9 (11), 8722-8747. 2009

[8] J. Lloret, M. Garcia, A. J. Ferrer, M. Morant, HDTV compression analysis for storage and transmission over Internet, The 5th WSEAS International Conference on DATA NETWORKS, COMMUNICATIONS and COMPUTERS (DNCOCO 2006), Bucarest (Rumania), 16-18 de Octubre de 2006.

[9] J. Lloret, M. Garcia, A. J. Ferrer, M. Morant, Editing and compressing HDTV videos for storage and transmission over Internet, WSEAS Transactions on Signal Processing, Issue 10, Vol. 2, pp. 1379-1386, Octubre 2006.

[10] M. Garcia, J. Lloret, M. Edo, R. Lacuesta, IPTV Distribution Network Access System Using WiMAX and WLAN Technologies, Int. Symposium on High Performance Distributed Computing (HPDC 2009), Munich (Alemania), 11-13 de Junio 2009.

[11] M. Atenas, M. Garcia, A. Canovas, J. Lloret, A MPEG-2/MPEG-4 Quantizer to Improve the Video Quality in IPTV services, The Sixth International Conference on Networking and Services (ICNS 2010), Cancun (Mexico), 7-13 de marzo de 2010.

[12] A. Canovas, M. Garcia, J. Lloret, M. Atenas, R. Rizo, Improving IPTV QoE taking the suitable MPEG-2/MPEG-4 Quantizer based on Jitter, Delay and lost packets measurements, Int. Journal On Advances in Telecommunications, Vol. 3, Issue: 3&4, pp. 129 - 139, Dic. 2010.

[13] A. Canovas, M. Garcia, J. Lloret, J. Tomás, 3D IPTV Study and Performance, The Second International Conference on Evolving Internet (INTERNET 2010), Valencia, España, 20-24 de Sep. de 2010.

[14] M. Atenas, S. Sendra, M. Garcia, J. Lloret, IPTV Performance in IEEE 802.11n WLANs, IEEE Global Communications Conference (IEEE Globecom 2010), Miami (USA), 6 - 10 de Diciembre de 2010.

[15] J. Lloret, M. Atenas, A. Canovas, M. Garcia, A Network Management Algorithm Based on 3D Coding Techniques for Stereoscopic IPTV Delivery, IEEE Global Communications Conference (IEEE Globecom 2011), Houston, Texas, USA, 5 - 9 de Diciembre de 2011.

[16] C. Barambones, D. Pascual, J. R. Diaz, J. Lloret, Una Nueva Herramienta para Testear el Rendimiento de una Red IP, XI Jornadas de Ingeniería Telemática (JITEL 2013), Granada (España), 28-30 de Octubre de 2013.

[17] A. S. Sadiq, K. A. Bakar, K. Z. Ghafoor, J. Lloret, An Intelligent Vertical Handover Scheme for Audio and Video Streaming in Heterogeneous Vehicular Networks, Mobile Networks & Applications, Vol. 18, Issue 6, pp. 879-895, Dec. 2013.

[18] L. D. P. Mendes, J. J. P. C. Rodrigues, J. Lloret, S. Sendra, Cross-layer Dynamic Admission Control for Cloud-based Multimedia Sensor Networks, IEEE Systems Journal, Vol. 8, Issue 1, pp. 235-246, Febrero 2014.

[19] J. R. Diaz, J. Lloret, J. M. Jiménez, M. Hammoui, A New Multimedia-Oriented Architecture and Protocol for Wireless Ad Hoc Networks, International Journal of Ad Hoc and Ubiquitous Computing, Vol. 16, Issue: 1, pp. 14-25, 2014.

[20] A. López-Herreros, A. Canovas, J. M. Jimenez, J. Lloret, A New IP Video Delivery System for Heterogeneous Networks using HTML5, IEEE Global Communications Conference (IEEE Globecom 2015), London (UK), 8-12 de Junio de 2015.

[21] M. García, A. Canovas, M. Edo, J. Lloret, A QoE Management System for Ubiquitous IPTV Devices, The third International Conference on Mobile Ubiquitous Computing, Systems, Services and technologies (UBICOMM 2009), Sliema (Malta), 11-16 de Octubre de 2009.

[22] M. Atenas, M. Garcia, A. Canovas, J. Lloret, IPTV Transcoding to Avoid Network Congestion, 6th Int. Conference on Networking and Services (ICNS 2010), Cancun (Mexico), 7-13 de marzo de 2010.

[23] J. Lloret, M. Garcia, M. Atenas, A. Canovas, A QoE management system to improve the IPTV network, International Journal of Communication Systems, Vol. 24, Issue: 1, pp. 118-138, January 2011.

[24] J. Lloret, M. Garcia, M. Atenas, A. Canovas, A Stereoscopic Video Transmission Algorithm for an IPTV Network Based on Empirical Data, International Journal of Communication Systems, Vol. 24, Issue: 10, pp. 1298-1329, Octubre 2011.

[25] A. Canovas, D. Bri, S. Sendra, J. Lloret, Vertical WLAN Handover Algorithm and Protocol to Improve the IPTV QoS of the End User, IEEE International Conference on Communications (ICC 2012), Ottawa (Canada), 10-15 de Junio de 2012.

- [26] J. Lloret, A. Canovas, J. J. P. C. Rodrigues, K. Lin, A Network Algorithm for 3D/2D IPTV Distribution using WiMAX and WLAN Technologies, *Multimedia Tools and Applications*, Vol. 67, Issue 1, pp. 7-30, Noviembre 2013.
- [27] J. Lloret, A. Canovas, J. Tomas, M. Atenas, A Network Management Algorithm and Protocol for Improving QoE in Mobile IPTV, *Computer Communications*, Vol. 35, Issue 15, pp. 1855-1870, Septiembre 2012.
- [28] J. R. Diaz, J. Lloret, J. M. Jimenez, S. Sendra, MWAHCA: A Multimedia Wireless Ad Hoc Cluster Architecture, *Scientific World Journal*, Vol. 2014, Issue: Article ID 913046, 14 pages, March 2014.
- [29] J. R. Diaz, J. Lloret, J. M. Jimenez, J. J. P. C. Rodrigues, A QoS-Based Wireless Multimedia Sensor Cluster Protocol, *International Journal of Distributed Sensor Networks*, Vol. 2014, Issue: Article ID 480372, páginas 17, Febrero 2014.
- [30] J. R. Diaz, J. Lloret, J. M. Jimenez, S. Sendra, Joel J. P. C. Rodrigues, Fault Tolerant Mechanism for Multimedia Flows in Wireless Ad Hoc Networks based in Fast Switching Paths, *Mathematical Problems in Engineering*, Vol. 2014, Issue: Article ID 361543, páginas 12, 2014.
- [31] J. M. Jiménez, J. R. Diaz, S. Sendra, J. Lloret, Choosing the Best Video Compression Codec Depending on the Recorded Environment, *IEEE Global Communications Conference (IEEE Globecom 2014)*, Austin, Texas (USA), 8-12 de Diciembre de 2014.
- [32] J. M. Jimenez, A. Canovas, A. López-Herreros, J. Lloret, A New Algorithm to Improve the QoE of IPTV Service Customers, *IEEE Global Communications Conference (IEEE Globecom 2015)*, London (UK), 8-12 de Junio de 2015

IV Jornadas de Innovación Educativa en Ingeniería Telemática (JIE)

Evaluación para el aprendizaje aplicando revisión entre iguales: caso de estudio en una asignatura de programación

Raquel M. Crespo-García, Iria Estévez-Ayres, Carlos Delgado Kloos

Departamento de Ingeniería Telemática,

Universidad Carlos III de Madrid

Avda de la Universidad, 30, E-28911 Leganés (Madrid) Spain.

{rcrespo, ayres, cdk}@it.uc3m.es

Resumen—En este artículo se presenta una experiencia de evaluación para el aprendizaje, aplicando la metodología de revisión entre iguales en la asignatura Arquitectura de Sistemas del segundo curso del Grado en Ingeniería de Tecnologías de las Telecomunicaciones. La actividad propuesta tiene como objetivo prioritario fomentar el aprendizaje reflexivo y que los alumnos afiancen los fundamentos teóricos de la materia. Asimismo, desarrolla competencias transversales clave para el futuro profesional de un ingeniero. El artículo describe la actividad desarrollada por los alumnos, así como su motivación, y presenta resultados preliminares muy prometedores, basados en encuestas de los alumnos, así como una comparativa respecto a ediciones anteriores.

Palabras Clave—Evaluación para el aprendizaje, revisión entre iguales, evaluación formativa, programación.

I. INTRODUCCIÓN Y MOTIVACIÓN

La implantación de los nuevos planes de estudio adaptados al Espacio Europeo de Educación Superior (Declaración de Bolonia [1]) se ha traducido en las universidades españolas en un modelo basado en evaluación continua, con pruebas distribuidas a lo largo de todo el cuatrimestre, que fuerza a los estudiantes a realizar continuamente distintos tipos de evaluaciones. Sin embargo, el mero aumento de las pruebas de evaluación no implica automáticamente un aumento en la implicación y aprendizaje de los alumnos [2].

Numerosos estudios recogen el impacto de la evaluación en el aprendizaje. Qué se evalúa determina en gran medida qué aprenden los alumnos. Así, una estrategia habitual en nuestros alumnos es centrarse en las pruebas que deben realizar y estudiar únicamente los conceptos necesarios para dichas pruebas. Hemos observado en los últimos años que en aquellas asignaturas eminentemente prácticas una gran mayoría de los alumnos intentan dominar aquello que se les va a preguntar en el examen sin profundizar en la teoría o memorizándola sin llegar a entenderla.

La *evaluación para el aprendizaje* (en contraposición a la tradicional evaluación del aprendizaje) persigue no sólo verificar los conocimientos y competencias adquiridos por los alumnos sino también ayudar a determinar los siguientes pasos en el proceso de aprendizaje. Al objetivo meramente calificador suma un propósito formativo. Las principales estrategias de evaluación para el aprendizaje —alinear objetivos de aprendizaje, realimentación formativa, auto- y co-evaluación y el uso formativo de las pruebas de calificación— han

demostrado un claro efecto positivo en la mejora del proceso de aprendizaje [3].

Asumiendo que los alumnos desarrollan una estrategia de aprendizaje focalizada en la evaluación (lo que *in extremis* puede incluso derivar en que el aprendizaje en sí sea un efecto colateral en lugar del objetivo prioritario), la evaluación para el aprendizaje busca alinear los objetivos de aprendizaje y aprovechar el potencial formativo de las pruebas de evaluación.

Black y Wiliam [4] destacan los siguientes factores asociados a la mejora del aprendizaje mediante la evaluación:

1. Proporcionar realimentación efectiva a los alumnos.
2. Que los alumnos se involucren activamente en su propio aprendizaje.
3. Adaptar la docencia de acuerdo a los resultados de la evaluación.
4. Reconocer la profunda influencia que tiene la evaluación en la motivación y autoestima de los alumnos, elementos ambos con influencia clave en el aprendizaje.
5. La necesidad de los estudiantes de ser capaces de evaluarse a sí mismos y entender cómo mejorar.

Por otra parte, el hecho de poder profundizar en los conceptos e ideas y poder realizar críticas constructivas del trabajo de otros es una habilidad clave de cara al futuro profesional de los alumnos. De hecho, el currículum de Computer Engineering de la ACM [5], así lo recoge, destacando (entre otras) las siguientes capacidades:

1. Comunicar ideas de forma escrita de manera efectiva.
2. Comprender y ofrecer críticas constructivas a las presentaciones de otros.
3. Defender (de forma educada pero efectiva) una posición.

Para fomentar el pensamiento crítico y, simultáneamente, facilitar que los estudiantes interioricen los conceptos teóricos necesarios en una asignatura de programación, se ha propuesto a los alumnos una actividad opcional en la que se aplica la metodología de revisión entre iguales.

En el ámbito educativo, se define la revisión entre iguales como un proceso en el que los alumnos evalúan la cantidad, nivel, valor, calidad o éxito de los productos o resultados de aprendizaje de compañeros de similar nivel [6]. Topping [6] realiza un completo compendio de los beneficios asociados al uso de esta metodología en educación superior. En línea con

los objetivos planteados previamente, la revisión entre iguales fomenta el aprendizaje reflexivo, desarrolla el pensamiento crítico y las capacidades de evaluación de los alumnos, lo cual se suma a las competencias relacionadas con la creación en sí del producto a revisar (*Evaluación y Síntesis* se sitúan en los niveles superiores de la taxonomía de Bloom [7]).

En concreto, en este caso los alumnos debían:

1. Redactar una pregunta sobre los contenidos teóricos de la asignatura. En este punto se intenta reforzar la habilidad 1 del listado anterior.
2. Revisar el trabajo de varios compañeros y realizar una crítica constructiva del mismo, proporcionándoles realimentación (habilidad 2).
3. Reflexionar de forma crítica sobre el propio trabajo y la realimentación recibida (habilidades 1 y 3).

La experiencia propuesta busca, en definitiva, potenciar el potencial formativo de una actividad de evaluación, en línea con las estrategias de *evaluación para el aprendizaje*, para trabajar capacidades y competencias de los alumnos críticas no sólo para su éxito en la asignatura sino en su futuro profesional.

Este artículo se organiza como se especifica a continuación. Primero se describe la asignatura en que se ha realizado la experiencia (sección II) y la actividad realizada (sección III). A continuación se presentan los resultados obtenidos (sección IV). Finalmente, la sección V resume las principales contribuciones y conclusiones de este trabajo.

II. CONTEXTO

La asignatura Arquitectura de Sistemas (AS) se imparte en el segundo año del Grado en Tecnologías de las Telecomunicaciones en la Universidad Carlos III de Madrid, tanto para la titulación bilingüe como para la titulación totalmente en castellano. La edición de 2014-15 (el año de esta experiencia) es la cuarta de esta asignatura en esta titulación. El número total de alumnos de esta asignatura oscila entre 90 y 120, en concreto, en la edición de 2014-15, el grupo en castellano (CAST) contaba con 73 alumnos, mientras que el grupo bilingüe (BIL) lo conformaban 19.

La metodología docente de la asignatura combina clases *magistrales* en grupos grandes (CAST-M y BIL-M) y sesiones prácticas en laboratorios, con grupos relativamente reducidos de aproximadamente 40 alumnos (CAST-R1, CAST-R2, BIL-R). En la edición 2014-15, la misma docente se ha hecho cargo de casi todos los grupos, exceptuando el grupo BIL-R, que se impartió de forma compartida con otro docente.

Debido a que el grupo de alumnos bilingüe tiene un perfil muy distinto del grupo en castellano y a que el profesorado era diferente (todos los grupos de castellano impartidos por el mismo docente, mientras que el grupo bilingüe fue impartido por 2 docentes distintos), lo cual hace que los resultados sean difícilmente comparables, se decidió dejar el grupo bilingüe fuera del estudio presentado en este artículo.

La metodología docente gira en torno al aprendizaje activo [8], con una serie de prácticas y proyectos en equipo de dificultad creciente, combinados con exámenes individuales a lo largo del cuatrimestre. Existe un examen final, opcional, dirigido a los alumnos que no sigan la evaluación continua, pero los alumnos nunca han optado por dicho examen en

ninguna de las ediciones de la asignatura. Esto es en gran parte debido a que la política de evaluación de la Universidad Carlos III de Madrid incentiva seguir la evaluación continua, pues si la evaluación de una asignatura es continua al 100 % (como en este caso), el examen final de la convocatoria ordinaria puntúa sobre 6 puntos (y la puntuación necesaria para aprobar es un 5 sobre 6, o lo que es lo mismo 8,33 sobre 10).

El temario de la asignatura abarca la enseñanza del lenguaje de programación C, así como conceptos de sistemas operativos, procesos, hilos, comunicación, concurrencia y diferentes mecanismos de sincronización. Estos últimos temas se abordan tanto desde un punto de vista más teórico, usando pseudo-código, como desde un punto de vista más práctico, usando tanto el lenguaje C como el lenguaje Java y sus distintos mecanismos de sincronización y comunicación entre procesos e hilos. La asignatura tiene, por tanto, una componente importante de programación, tanto en C como en Java, que se refleja en los proyectos y prácticas a realizar por los alumnos.

Los alumnos han cursado previamente dos asignaturas de programación, durante el primer curso:

1. Programación: introducción a la programación (Java)
2. Programación de Sistemas: orientación a objetos, estructuras de datos y algoritmos (Java)

En Arquitectura de Sistemas los alumnos se enfrentan a un nuevo lenguaje, C, que al ser de más bajo nivel permite introducir los conceptos de arquitectura de sistema subyacentes. Durante sus dos primeras ediciones, el temario de la asignatura se limitaba a la enseñanza de dicho lenguaje [9]. El resto del temario se incorporó a partir de la tercera edición, para paliar las carencias detectadas en el plan de estudios y que afectaban a asignaturas de cursos superiores.

La asignatura plantea como objetivos específicos:

1. diseñar y desarrollar aplicaciones concurrentes en el lenguaje de programación C;
2. saber identificar la concurrencia necesaria en una aplicación, detectar y prevenir situaciones problemáticas, usando los mecanismos de sincronización adecuados para cada caso;
3. aplicar técnicas de trabajo en equipo para desarrollar una aplicación compleja;
4. usar herramientas de desarrollo de aplicaciones susceptibles de aplicación en el entorno profesional (como el controlador de versiones Subversion¹, máquinas virtuales y el conjunto de herramientas Valgrind²);
5. desarrollar técnicas de autoaprendizaje.

El esquema de evaluación continua aplicado hace que los alumnos presten especial atención a los aspectos prácticos, lo cual es lógico en una asignatura de programación. Sin embargo, se evidencian a menudo carencias teóricas que dificultan el aprendizaje en profundidad de la materia. Por ello, en la presente edición se intentaron afianzar los fundamentos teóricos, proponiendo a los alumnos actividades opcionales (ver sección III). Como objetivo adicional, estas actividades opcionales se planteaban al inicio del curso, cuando la carga lectiva de prácticas era menor, para fomentar que los alumnos

¹<http://subversion.apache.org>

²<http://valgrind.org>

Has realizado (y completado) la actividad opcional*

- Sí
 No

Si no la has realizado (o no la has completado), por favor, indica por qué

Si la has intentado, indica si te ha parecido útil

1 2 3 4 5

Nada útil Muy útil

Con respecto a la pregunta anterior, ¿por qué?

Puedes incluir aquí aspectos negativos o positivos de la actividad o de la plataforma

Según tu experiencia, merece la pena implantar esta actividad en futuras ediciones*

- Sí
 No
 No lo sé

Figura 1. Encuesta sobre la actividad.

trabajasen la asignatura de forma continua y desde el primer momento.

La asignatura Arquitectura de Sistemas aplica una metodología de mejora continua que, entre otras cosas, incluye encuestas periódicas a los alumnos sobre la evolución del curso [10]. En concreto, se plantean tres encuestas, voluntarias y anónimas, a lo largo del cuatrimestre (además de las oficiales de la universidad):

1. Al final del primer mes
2. Al final del segundo mes
3. Al final del curso

Los resultados de estas encuestas internas se evalúan en el momento, durante el curso, permitiendo aplicar las medidas necesarias para mejorar la metodología docente, subsanar posibles deficiencias y facilitar el aprendizaje de los alumnos.

III. METODOLOGÍA

III-A. Descripción de la actividad

Para fomentar el aprendizaje reflexivo, se planteó a los alumnos una actividad opcional consistente en la creación de forma colaborativa de preguntas tipo test. Con vistas a que los alumnos tuvieran la oportunidad de reflexionar sobre otros aspectos del tema, no sólo debían proponer su propia pregunta sino revisar y evaluar preguntas propuestas por sus

compañeros y, en su caso, proponer correcciones a las mismas. En concreto, se aplicó la metodología de revisión entre iguales, siguiendo un proceso estructurado en las siguientes etapas:

1. Fase 1: creación de preguntas.

Cada alumno debía proponer una pregunta tipo test, incluyendo enunciado y opciones, y especificar cuál era la respuesta correcta. Para esta fase disponían de un plazo de 5 días.

2. Fase 2: revisión de preguntas de compañeros.

A cada alumno se le asignaban tres preguntas propuestas por sus compañeros. El alumno veía por una parte el enunciado con las opciones y, por otra, la solución indicada por el autor. Los revisores debían comprobar la pregunta, verificando que se ajustase al tema, se comprendiera correctamente, no existieran problemas de ambigüedad, etc. Además, debían indicar cuál era la respuesta correcta, a su juicio y justificar las posibles discrepancias si su respuesta no coincidía con la del autor. Finalmente, podían sugerir correcciones o incluso una redacción alternativa para mejorar la pregunta.

Además de la realimentación cualitativa, los revisores también debían incluir una valoración cuantitativa de la pregunta, evaluando aspectos como la adecuación al

Cuadro I
PARTICIPACIÓN.

	Participación	Participación (%)
Completa	45	61,64 %
Parcial	10	13,70 %
No participa	18	24,66 %
Total	73	100,00 %

tema, dificultad y, en general, calidad de la pregunta. Asimismo, se les pedía que ordenasen comparativamente las tres preguntas asignadas para revisión.

Esta fase tenía una duración de 4 días.

3. Fase 3: mejora de preguntas.

En la fase final, los autores recibían los comentarios y valoraciones de sus compañeros y podían proponer una versión modificada de su pregunta, de acuerdo a los comentarios recibidos. Esta fase tenía una duración de 4 días.

III-B. Metodología

Para participar en la experiencia, los alumnos se dividieron en dos grupos, coincidentes con los grupos administrativos (con lo que puede asumirse asignación aleatoria). Cada grupo realizaba exactamente el mismo proceso, salvo que cada uno trabajó en temas distintos.

Si bien se trataba de una actividad opcional, los docentes estimaron justo recompensar el esfuerzo dedicado por los alumnos con una pequeña puntuación. En concreto, los alumnos podían sumar hasta 0,25 puntos en su calificación:

- 0,1 puntos simplemente por completar todas las fases del proceso (es decir, por participar);
- 0,15 puntos dependiendo de su rendimiento en dicha actividad.

Al finalizar la actividad, se pidió a los alumnos que completasen una encuesta sobre la misma. Dicha encuesta formaba parte de las encuestas de mejora continua de la calidad ya establecidas previamente en la asignatura (ver sección II).

La figura 1 muestra las preguntas planteadas en la encuesta sobre la actividad. En la siguiente sección se analizan los resultados de la misma.

IV. RESULTADOS

IV-A. Participación

La mayoría de los alumnos (75,33 %) participaron en la actividad. Sólo un porcentaje pequeño de los participantes no llegó a completar las tres fases. La causa de ello, según las respuestas de los alumnos en la encuesta, es que se les pasó el plazo de entrega de alguna fase. La figura 2 muestra los datos de participación en la actividad, detallados en la tabla I.

IV-B. Resultados cuantitativos

De los 55 alumnos participantes en la experiencia, respondieron la encuesta sobre la actividad opcional 22 (40 %). Pese a tratarse de una encuesta voluntaria, creemos que este porcentaje es suficientemente significativo para concluir que los resultados son representativos de la opinión de los alumnos.

La figura 3 muestra la distribución de respuestas de la pregunta "Si la has intentado, indica si te ha parecido útil".

Participación en actividad

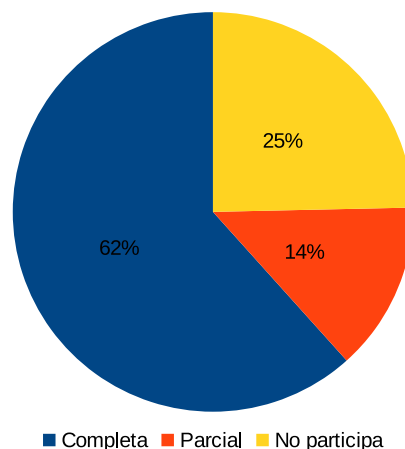


Figura 2. Participación en la actividad.

Como puede verse en la figura, los alumnos consideraron mayoritariamente útil la actividad.

Utilidad de la actividad

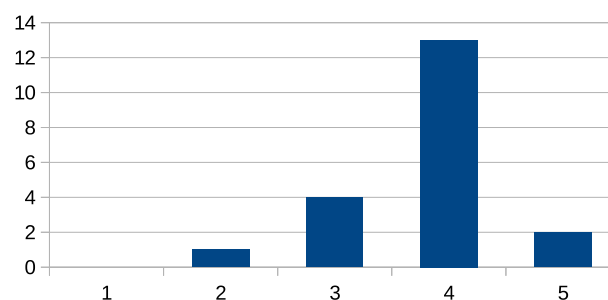


Figura 3. Utilidad de la actividad.

Lo que es aún más revelador, el 95 % de los alumnos responde positivamente la pregunta "Según tu experiencia, merece la pena implantar esta actividad en futuras ediciones". Sólo un alumno (5 %) responde "No lo sé" y ninguno de forma negativa.

IV-C. Opiniones cualitativas

En la encuesta realizada a los alumnos, se les preguntaba sobre su opinión acerca de la actividad (figura 4). En la tabla II se muestra el análisis realizado a estos comentarios. Como se puede observar, en general, los alumnos destacan la utilidad de la actividad en su proceso de estudio, para afianzar conceptos y entender en profundidad la materia, así como para llevar la materia al día desde el primer momento. Algunos comentarios en esta línea:

"Lógicamente para hacer la pregunta tienes que haberte leído por lo menos la materia de la asignatura y, al corregir, te sirve para saber materia que no sabías."

"Es una manera al principio de la asignatura de ir al día. Todos sabemos que las primeras semanas

IV-D. Resultados académicos

En la tabla III se muestra la comparativa de los resultados académicos entre esta edición y la anterior. Se puede apreciar que, en el caso del grupo en castellano, el porcentaje de alumnos que siguieron la evaluación continua es similar (aprox. un 79%). Sin embargo, en esta edición, el porcentaje de aprobados en el grupo de castellano (en el que se llevó a cabo esta experiencia) sobre los que siguieron la continua ha aumentado un 10% respecto al año anterior.

Si analizamos para el conjunto de los aprobados la participación en esta experiencia, se puede observar que el 84,21% de los aprobados de la asignatura participaron en la misma (ver tabla IV).

En el grupo bilingüe los resultados son notablemente inferiores en esta edición. Sin embargo, al tratarse de un grupo con un número muy reducido de alumnos con un perfil muy concreto, no pueden considerarse significativas estas cifras, pues la variación en el perfil es demasiado alta entre edición y edición.

V. CONCLUSIONES

En este artículo se ha presentado una experiencia desarrollada en la asignatura Arquitectura de Sistemas, del Grado en Ingeniería de Tecnologías de las Telecomunicaciones. El objetivo, en línea con las estrategias de *evaluación para el aprendizaje*, es potenciar el potencial formativo de una actividad de evaluación, para trabajar capacidades y competencias de los alumnos críticas no sólo para su éxito en la asignatura sino en su futuro profesional. Para ello se ha aplicado la metodología de revisión entre iguales para desarrollar habilidades transversales como el pensamiento crítico y la capacidad de evaluación, fomentar el aprendizaje reflexivo y afianzar los conocimientos teóricos de los alumnos.

La comparativa de los resultados académicos de la edición en que se ha desarrollado la actividad muestra una mejora en el porcentaje de alumnos que ha superado la asignatura respecto a la edición anterior (sin la actividad). Los resultados de las encuestas realizadas a los alumnos confirman la utilidad de la actividad, así como su efecto positivo en otros aspectos como la motivación de los alumnos.

Si bien los resultados obtenidos son positivos, se plantea como trabajos futuros un análisis en profundidad de los datos académicos para confirmar el efecto positivo de la actividad en el aprendizaje.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por la Comunidad de Madrid a través de la Red de Excelencia eMadrid (S2013/ICE-2715) y por el Ministerio de Economía y Competitividad a través de los proyectos EEE (TIN2011-28308-C03-01) y RESET-UC3M: REFORMULANDO ECOSISTEMAS ESCALABLES EDUCATIVOS (TIN-2014-53199-C3-1-R).

REFERENCIAS

- [1] Declaración de Bolonia, "El espacio europeo de la enseñanza superior," Declaración conjunta de los ministros europeos de educación reunidos en Bolonia el 19 de junio de 1999.
- [2] P. Surgenor, "Teaching toolkit. effect of assessment on learning," UCD Teaching and Learning Resources, 1 2010. [Online]. Available: www.ucd.ie/t4cms/UCDTLT0031.pdf

- [3] M. T. Flórez and P. Sammons, "Assessment for learning: effects and impact." Oxford University Department of Education, Tech. Rep., 2013. [Online]. Available: <http://cdn.cfbt.com/~/media/cfbtcorporate/files/research/2013/r-assessment-for-learning-2013.pdf>
- [4] P. Black and D. Wiliam, *Inside the black box: Raising standards through classroom assessment*. Granada Learning, 1998.
- [5] The Joint Task Force on Computing Curricula, "Computer Engineering 2004. Curriculum Guidelines for Undergraduate Degree Programs in Computer Engineering. A Report in the Computing Curricula Series," IEEE Computer Society. Association for Computing Machinery, Tech. Rep., 2004. [Online]. Available: https://www.acm.org/education/education/curric_vols/CE-Final-Report.pdf
- [6] K. Topping, "Peer assessment between students in colleges and universities," *Review of Educational Research*, vol. 68, no. 3, pp. 249–276, Fall 1998.
- [7] B. S. Bloom, *Taxonomy of educational objectives: The classification of educational goals. Handbook I, cognitive domain*. Longmans, Green, 1956.
- [8] I. Estévez-Ayres, C. Alario-hoyos, M. Pérez-Sanagustín, R. M. Crespo-García, D. Leony, and H. A. Parada G, "Validación por la Comunidad Docente de una Metodología de Aprendizaje Activo para Cursos de Programación," in *Jornadas de Innovación Educativa en Ingeniería Telemática (JIE)*, Granada, Oct. 2013.
- [9] A. Pardo, I. Estévez-Ayres, P. Basanta-Val, and D. Fuentes-Lorenzo, "Programación en C con aprendizaje activo, evaluación continua y trabajo en equipo: caso de estudio," in *XVI Jornadas de Enseñanza Universitaria de la Informática (JENUI 2010)*, Julio 7-9 2010.
- [10] I. Estévez-Ayres, C. Alario-Hoyos, M. Pérez-Sanagustín, A. Pardo, R. M. Crespo-García, D. Leony, H. A. Parada G., and C. Delgado-Kloos, "A methodology for improving active learning engineering courses with a large number of students and teachers through feedback gathering and iterative refinement," *International Journal of Technology and Design Education*, vol. 25, no. 3, pp. 387–408, August 2015. [Online]. Available: <http://link.springer.com/10.1007/s10798-014-9288-6>

Coevaluación de trabajos mediante el empleo de una aplicación móvil

José Luis Tornos¹, José Luis Salazar¹, Guillermo Azuara¹, M^a Gloria Mendieta y Pilar Del Valle

Departamento Ingeniería y Electrónica de Comunicaciones

Universidad de Zaragoza

C/ María de Luna 3.

{ jltornos, jsalazar, gazuara }@unizar.es; mgloria.mendieta@zoho.com; pdelvallegarcia@gmail.com

Resumen- En este artículo describimos el empleo de una aplicación móvil para la coevaluación de los alumnos en la asignatura de Comercio Electrónico. El uso de esta aplicación tiene dos objetivos primordiales: el primero de ellos es hacer partícipes a los alumnos en el proceso de evaluación de trabajos del resto de compañeros. El segundo es enlazar los contenidos de dos asignaturas del itinerario de Telemática del Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación (Comercio Electrónico y Seguridad en redes y servicios). De esta manera, los conceptos de seguridad, firma digital y anonimato, entre otros, encuentran una aplicación práctica en el empleo de una plataforma de recogida de información segura. Para lograr este objetivo se ha empleado una aplicación móvil que permite a los alumnos participar en el proceso de evaluación de los trabajos de sus compañeros.

Palabras Clave- coevaluación, aplicación móvil, recogida de información segura

I. INTRODUCCIÓN

La tarea de evaluación del aprendizaje por parte de los alumnos es una de las tareas más complejas del proceso educativo. Para intentar involucrar a los alumnos en los procesos de evaluación, y además intentar desarrollar competencias como la capacidad de autoevaluación y mejora del aprendizaje, desde la asignatura de Comercio Electrónico se ha optado por introducir la coevaluación como parte del mecanismo de evaluación de la asignatura.

La coevaluación forma parte de una de los denominados procedimientos de “evaluación alternativa” [1], cuyos principales objetivos son: contrastar respuestas, identificar disonancias, negociar significados...

Se tuvieron en cuenta diversos mecanismos para realizar la coevaluación de los trabajos presentados por los alumnos: formularios de papel, formularios online... pero finalmente se decidió emplear una plataforma de recogida de información segura desarrollada dentro del grupo de ingeniería telemática. De esta manera se introduce una herramienta TIC en el sistema de evaluación, tal y como se han desarrollado en trabajos previos [2].

La plataforma utilizada tiene la capacidad de ser empleada de diversas maneras, pudiendo adaptarse a diferentes procesos de recogida de información, variando tanto su formato como su nivel de seguridad. Se puede configurar de manera que no exista ningún control sobre la identificación de los participantes (sistema abierto), requerir un sistema de identificación login/password, o emplear un mecanismo de firmas digitales que permite realizar una

identificación del alumno como perteneciente a un grupo pero sin revelar su identidad.

Esta última opción fue la que finalmente se seleccionó y se puso en práctica ya que si se dota de anonimato a los participantes en cualquier proceso de evaluación o valoración, éstos se sienten más libres a la hora de expresarse [3] que si se pudiese establecer cualquier tipo de conexión entre la evaluación y el evaluador.

De esta manera, se conseguiría mostrar a los alumnos una aplicación práctica de un sistema, orientado originalmente hacia el eVoting, que emplea firmas en anillo y permite una identificación anónima de los participantes. Estos conceptos, relacionados con la criptografía, las firmas digitales y la seguridad de la información, se desarrollan en la asignatura “Seguridad en redes y servicios”, perteneciente al mismo itinerario de telemática que la asignatura de “Comercio Electrónico”.

Un motivo añadido por el que se decidió emplear esta herramienta es que de esta manera se introducía un concepto TIC, como es una aplicación móvil, desarrollada por el departamento, que facilita la participación de los alumnos en los procesos de evaluación [4].

Otro de los objetivos de la implantación de este sistema de coevaluación [5] es incentivar en cierta medida la asistencia y participación de los alumnos en las sesiones de presentación de los trabajos de sus compañeros. Para intentar lograr esto, únicamente se permitirá participar en la coevaluación a aquellos alumnos que hayan asistido a clase, algo que parece obvio, pero que al emplear un sistema de identificación anónima como el desarrollado, exige la elaboración de un censo de participantes al inicio de cada sesión de presentación de trabajos.

El resto del artículo está organizado de la siguiente manera. En la sección II se realiza una descripción de la plataforma empleada. En la sección III se explica la metodología empleada para la consecución de la propuesta de innovación docente. En la sección IV se muestran y analizan los resultados obtenidos. Finalmente, en la sección V, se indican las conclusiones y líneas futuras de trabajo.

II. DESCRIPCIÓN DE LA PLATAFORMA

La plataforma empleada para realizar la recogida de información está basada en el trabajo descrito en [6]. Dicha plataforma tiene unas características que la hacen propicia para ser empleada en los métodos de coevaluación:

1. La plataforma permite el uso de diferentes mecanismos de identificación de los participantes: abierto (sin identificación); *login/password*; o mediante el uso de firmas digitales.
2. La plataforma es multiplataforma, es decir, se pueden emplear un navegador web (Firefox) o bien una aplicación móvil específica (Android).
3. El diseño de la plataforma permite que la recogida de la información sea fácilmente adaptable a las necesidades de cada uso. Permite insertar formularios, vídeos u otros archivos multimedia.

La plataforma se compone de dos módulos diferenciados: cliente y servidor. De esta manera se consigue independizar los requisitos que se solicitará al usuario (cliente) de lo desarrollado en la parte del servidor, encargado de la gestión de las votaciones, control de acceso y gestión de la base de datos con la información recogida.

Para poder participar en los procesos de valoración, los usuarios pueden acceder desde cualquier navegador web, cuando se emplea un mecanismo de identificación abierto (sin identificación de usuario) o un sistema de *login/password*. Cuando el sistema emplea firmas digitales, es necesario que los usuarios descarguen una extensión específica del navegador *Firefox* desarrollada para gestionar las firmas en anillo, o bien una aplicación *Android* desarrollada para acceder desde terminales con dicho sistemas operativo. La descarga de la extensión del navegador se realiza desde la misma página del gestor de votaciones mientras que la aplicación para el móvil está disponible en Google Play Store [7].

De manera más específica, el empleo del sistema de identificación basado en firmas digitales permite que la identificación se realice de manera anónima. Esto se consigue mediante el uso de firmas en anillo, que permiten identificar a un participante como parte de un censo pero sin especificar qué usuario es en concreto.

Además, se ha implementado el sistema de manera que acompañando a cada firma se incluya un campo, *linking tag*, dependiente tanto de la clave privada de los participantes como de los parámetros de una votación concreta, que permite realizar un control sobre los usuarios participantes.

Este *linking tag* también puede ser empleado para conocer la evolución de las valoraciones de los usuarios a lo largo del tiempo (en caso de que una valoración pueda contener más de una ronda); o bien a lo largo de distintas valoraciones (siempre y cuando se mantengan los parámetros iniciales).

III. METODOLOGÍA DE TRABAJO

La metodología de coevaluación se va a llevar a cabo en la asignatura de Comercio Electrónico. Esta asignatura se enmarca como obligatoria dentro de itinerario de Telemática del Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación y optativa en el Grado de Ingeniería Informática.

La planificación de la asignatura se divide en dos grandes apartados: teoría y realización de un proyecto web que

incluya un negocio virtual. La explicación y desarrollo de la parte teórica se lleva a cabo mediante clases magistrales y se evalúa al finalizar el cuatrimestre mediante un examen escrito de preguntas tipo test. El desarrollo de una tienda virtual se realiza individualmente o en pareja (a elección de los alumnos). Este trabajo se implementa a través de diez sesiones de prácticas en las que en un par de ellas se incluye la realización de un plan de viabilidad. El objetivo principal de estas sesiones es hacer un análisis del negocio que se propone para su despliegue. Para ello es necesario hacer una introspectiva de los objetivos que se quieren alcanzar y resulta de gran utilidad analizar otros casos de uso exitosos, es decir, una empresa de Internet con cierto éxito avalado. El análisis de una empresa que opere en internet se evalúa mediante la exposición del trabajo al profesor y a sus compañeros, y es parte de la evaluación del desarrollo del negocio web.

La introducción del análisis de una empresa tiene un objetivo muy marcado: mostrar las diferentes oportunidades de negocio existentes en Internet. La elección de los alumnos del tipo de negocio que quieren desarrollar les permite exponer aquellas ideas que siempre quisieron mostrar en internet, pero que quizás nunca tuvieron tiempo de hacerlo. El único condicionante que se les exige es que sea un negocio. Y es aquí donde habitualmente flaqueaban sus desarrollos.

Era muy habitual que los grupos de trabajo implementaran tiendas con objetivos sociales o con productos marginales que usaban ellos en sus hobbies y con artificios multimedia muy espectaculares. No hay que olvidar que en el entorno educativo del que estamos hablando, el alumnado ya ha adquirido grandes capacidades de desarrollo para el material web y ven en el desarrollo de esta práctica una vía de expresión de sus inquietudes.

Este factor es sin duda ventajoso para incentivar el interés del alumnado en la asignatura, pero ofrece el peligro de que se queden en ese punto, sin avanzar en la importancia que tiene para el comercio electrónico su objetivo principal: obtener beneficios.

Para analizar el sistema de obtención de beneficios de empresas de Internet se eligieron empresas punteras nacionales e internacionales que abarcaran los típicos modelos de negocio: catálogo, servicios por suscripción, publicidad, contenidos digitales y modelos mixtos. Las empresas evaluadas fueron: Google, Zara, Yomvi, Marca, Blizzard, Amazon, Springer, Spotify, Netflix y Meetic.

Otro aspecto que también se tuvo en cuenta era la novedad de la medida y la falta de referencias del alumno a la hora de redactar el análisis. Así que el profesor hizo un ejemplo en clase a partir de un guion básico para la creación de un plan de viabilidad de una empresa nueva. Se añadieron datos históricos, tanto económicos, como de hitos importantes alcanzados por la empresa. Los datos bursátiles de capitalización y de recorrido histórico, sustituyeron a planes de marketing e inversiones iniciales. Siguiendo esta manera de trabajo se indujo a los alumnos para que hicieran lo mismo con las empresas que se les había asignado en suerte. El modelo ejemplar utilizado fue la empresa Facebook.

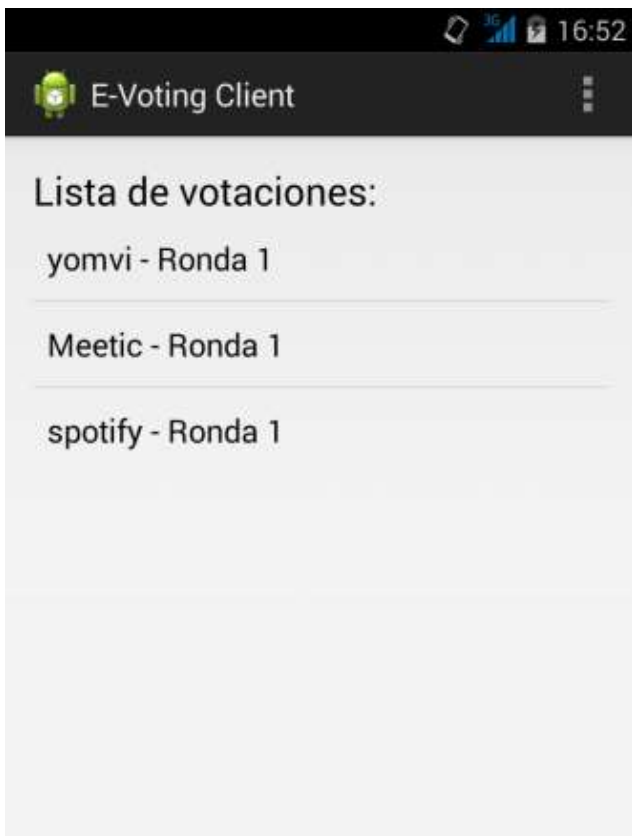


Fig. 1.- Pantalla de inicio de la aplicación móvil

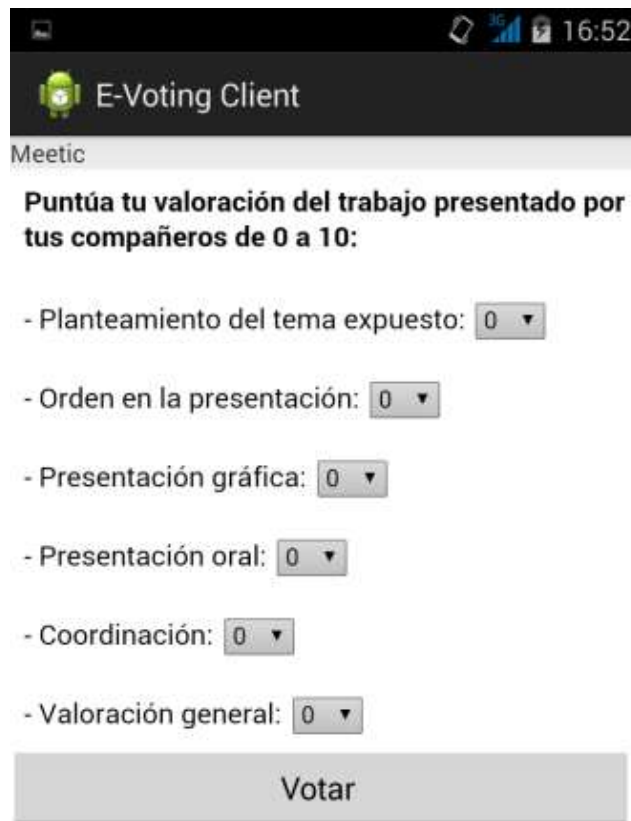


Fig. 2.- Pantalla de valoración de la aplicación móvil

Para involucrar a los alumnos en el desarrollo de la materia, se decidió realizar una coevaluación de los trabajos de análisis de negocios que operen en internet. De esta manera se estableció que cada grupo de 4 alumnos, desarrollase un trabajo de análisis sobre una compañía de entre las citadas anteriormente, disponiendo de 40 minutos por grupo para realizar la presentación dejando posteriormente un margen de 10 minutos para que se le realizasen preguntas aclaratorias. En la Fig. 1 se muestra cómo sería la pantalla de inicio de la aplicación para participar en la valoración de los trabajos expuestos.

Una vez distribuidos los alumnos en grupos se realizó una planificación de presentación de los trabajos por parte de los alumnos y se definió el sistema de evaluación del mismo. El proceso de evaluación se llevaría a cabo mediante un sistema de coevaluación, en el que la valoración del profesor contaría como una más entre todas las valoraciones recibidas. Únicamente podrían participar en el proceso de coevaluación aquellos alumnos que asistiesen a las presentaciones de los trabajos (excluyendo de dicha evaluación a los alumnos que realizasen la presentación). El profesor se reservaba el derecho de realizar “correcciones” sobre las valoraciones de los alumnos siempre que alguno de los aspectos valorados tuviesen una desviación mayor de 2 puntos, tanto hacia arriba como hacia abajo, respecto de la valoración del propio profesor.

Los aspectos a valorar, indicados a los alumnos de manera previa a la presentación de los trabajos, serían los siguientes:

- Planteamiento del tema expuesto (10%)
- Orden en la presentación (15%)
- Presentación gráfica (20%)
- Presentación oral (20%)
- Coordinación (10%)
- Valoración general (25%)

Entre paréntesis se indica el peso específico de cada uno de los aspectos a valorar. La información sobre dicha ponderación no fue facilitada a los alumnos. En la Fig. 2 se muestra la pantalla de valoración de la aplicación.

Para facilitar la participación de los alumnos en el proceso de valoración, se habilitó una semana de tiempo para que los asistentes valorasen el trabajo expuesto por sus compañeros. Se tomó esta decisión para permitir que alumnos que dispusieran de un móvil que emplease un sistema operativo distinto a Android, tuviesen un margen temporal suficiente para realizar la valoración. Además, de esta manera el profesor podía recordar a lo largo de las siguientes sesiones de la asignatura que aquellos alumnos que no hubiesen valorado el trabajo de sus compañeros todavía estaban a tiempo de participar.

El mecanismo de identificación para realizar las valoraciones que se implementó fue el de firmas en anillo (identificación anónima) para evitar cualquier tipo de sospecha por parte de los alumnos para que participasen en el proceso de evaluación.

TABLA 1

NÚMERO DE POTENCIALES EVALUADORES Y PARTICIPACIÓN

EVALUADORES POTENCIALES	PARTICIPANTES	PORCENTAJE
18	11	61%
22	10	45%
15	9	60%
17	8	47%
11	4	36%
14	4	29%
12	4	33%
12	3	25%
10	2	20%
10	3	30%

TABLA 2

NOTA MEDIA DE LOS TRABAJOS, NOTA DEL PROFESOR Y DIFERENCIA

MEDIA	NOTA PROFESOR	DIFERENCIA
7,25	7,9	-0,65
8,28	8,7	-0,42
7,81	7,1	0,71
7,85	7,6	0,25
7,8	7,05	0,75
7,68	7,8	-0,12
8,01	7,55	0,46
8	7,7	0,3
7,25	6,25	1
8,1	8,5	-0,4

Antes de iniciar cada sesión de presentación de los trabajos, el profesor elaboraba el censo de los alumnos asistentes y habilitaba el acceso a las valoraciones de cada uno de los grupos que presentasen en esa sesión, uno o dos trabajos distintos. Debido a que únicamente los alumnos que asistiesen a las sesiones de presentación de trabajos pueden formar parte del grupo de evaluadores, esta preparación previa debía realizarse al inicio de cada sesión, para así permitir que una vez terminada la presentación, los alumnos ya pudieran realizar la valoración del trabajo expuesto.

IV. EVALUACIÓN DE LOS RESULTADOS

La evaluación de los trabajos se llevó a cabo a lo largo del segundo cuatrimestre del curso 14-15, siendo la primera presentación el 24 de marzo y la última el 13 de mayo.

En la tabla 1 se muestran el número de asistentes a cada una de las diferentes sesiones, el número de participantes que finalmente participaron en la valoración de los trabajos y el porcentaje que representan. Como se observa en la tabla, la participación de los alumnos en el proceso de coevaluación fue bastante alto, superior al 60% en algunos casos.

El otro dato que también se aprecia es que conforme fueron avanzando las sesiones, el número de asistentes a las sesiones de presentación de trabajos se fue reduciendo, así como el número de participantes en el proceso de coevaluación, llegando a alcanzar en el peor de los casos, en la penúltima sesión, tan solo un 20% de participación.

En la tabla 2, se muestran las notas medias obtenidas por los distintos grupos, la nota que el profesor asignó a cada uno de dichos grupos y la diferencia entre la nota final y la nota del profesor. Podemos apreciar como en la mitad de los casos los alumnos valoraron el trabajo de sus compañeros con una nota superior a la que le asignó el profesor, mientras que en la otra mitad de los casos, fue el profesor el que valoró mejor que los alumnos el trabajo expuesto.

Otro de los datos a destacar es que en cinco de las diez presentaciones, hubo que hacer algún tipo de cambio en las valoraciones de los alumnos al exceder los dos puntos de diferencia las valoraciones respecto a las valoraciones del profesor. En total fueron 11 valoraciones individuales las que excedieron dicho límite, siendo cinco de dichas valoraciones, en alguno de los aspectos evaluados, inferiores en tres o más puntos a la valoración del profesor.

V. CONCLUSIONES

Dos de los principales objetivos que se plantearon a la hora de llevar a cabo el mecanismo de coevaluación se han conseguido. El primero, mostrar a los alumnos de la asignatura de Comercio Electrónico una aplicación práctica de conceptos desarrollados en la asignatura de Seguridad en redes y servicios, pertenecientes ambas al itinerario de telemática del Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación.

En segundo lugar, se ha involucrado a los alumnos en el proceso de evaluación del trabajo llevado a cabo por sus compañeros, haciéndolos partícipes en el proceso de evaluación.

Dada la variabilidad de resultados obtenidos, parece haber sido un gran acierto haber incluido varios campos de evaluación que analicen de forma metódica el trabajo valorado y no reducirlo todo a un valor que podría haber dado más variabilidad sobre la media de las apreciaciones de los alumnos.

A la vista de los resultados obtenidos, la participación de los alumnos en el proceso de coevaluación va decayendo conforme avanza la asignatura. Este hecho se puede deber a diversos factores: alumnos que abandonen la asignatura, desinterés en la participación de la coevaluación, falta de tiempo en las últimas fechas debido a la acumulación de trabajos...

Otra decisión sobre la que se debe reflexionar es si publicar las evaluaciones de cada trabajo, justo después de que se acabe el periodo de valoración, o esperar a la totalidad de las exposiciones para publicarlas. Se decidió por esta última opción para intentar eliminar el sesgo que podría ofrecer la valoración de alguien que haya sido evaluado de forma negativa, pensando, que de forma competitiva intentaría penalizar también al resto de sus compañeros. Pero también es precisamente, esa falta de competencia la que pudo haber incrementar la pérdida de interés por la coevaluación y la participación.

El corporativismo o animadversión personal es algo que puede influenciar negativamente en la coevaluación y para lo que hay que implementar medidas correctoras, como fue ignorar aquellas notas que se desviaban mucho de la apreciación del profesor y que, vistos los resultados, fueron de gran utilidad.

Sin embargo, esa falta de percepción de que la colaboración prestada por el alumno tuviera un reflejo inmediato, pudo haber sido clave para el descenso notable de participación. Esta sensación viene acusada por parte del profesor al recibir continuas preguntas sobre cuándo los alumnos conocerían la nota de dicha exposición.

Como continuación a la experiencia de coevaluación, se va a proponer en la misma asignatura una segunda propuesta de coevaluación. En esta segunda experiencia los alumnos presentarán, también por grupos, una tienda virtual que han ido desarrollando a lo largo del curso. La propuesta será similar a la primera, variando únicamente que se permitirá a los alumnos valorar su propio trabajo y además se premiará con medio punto más el trabajo mejor valorado y con un cuarto de punto más al segundo y tercer trabajo más valorado.

Con esta experiencia se esperan obtener datos para poder compararlos con los recogidos en la experiencia descrita en este artículo y ver cómo a un mismo grupo le afecta poder realizar una coevaluación y autoevaluación de su trabajo y el incentivo de poder aumentar la nota dependiendo de la valoración global recibida.

A lo largo de las dos experiencias se mantendrán los parámetros iniciales necesarios para llevar a cabo una votación con firmas en anillo para así poder comprobar, respetando siempre el anonimato de los participantes, si existe una gran diferencia entre el número de participantes en una y otra experiencia.

AGRADECIMIENTOS

Este artículo ha sido parcialmente financiado por el Proyecto UZ2014-TEC-02 (Universidad de Zaragoza) y por el grupo CeNITEQ (Communication Networks and Information Technologies for e-Health and Quality of Experience Group, Universidad de Zaragoza).

REFERENCIAS

- [1] Álvarez, I. (2008). Evaluación del aprendizaje: una Mirada retrospectiva y prospectiva desde la divulgación científica. *Revista Electrónica de Investigación Psicoeducativa*, 14(1), 235-272
- [2] Coll, C., Rochera, M. J., Mayordomo, R. y Naranjo, M. (2007). Evaluación continua y ayuda al aprendizaje. Una experiencia de innovación en Educación Superior con ayuda de TIC. *Electronic Journal of Research in Educational Psychology*, 5(3), 783-804
- [3] Palme J, Berglund M. Anonymity on the Internet. 2002. Available at: <http://people.dsv.su.se/~jpalme/society/anonymity.html> (Accedido 18-05-2015).
- [4] Keppell, M., Au, E. y Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment and Evaluation in Higher Education*, 31(4), 453-464.
- [5] Gessa Perera, Ana (2011). La coevaluación como metodología complementaria de la evaluación del aprendizaje. Análisis y reflexión en las aulas universitarias. *Revista de Educación*, 354, 749-764. http://www.revistaeducacion.educacion.es/re354/re354_30.pdf (Accedido 18-05-2015).
- [6] J. L. Tornos, J. L. Salazar, J. J. Piles, J. Saldana, L. Casadesus, J. Ruiz-Mas, J. Fernandez-Navajas: An eVoting System Based on Ring Signatures Network Protocols and Algorithms, vol. 6 n. 2 June, 2014. Pp. 38-54.
- [7] Aplicación móvil para la participación en procesos de recogida de información segura. <https://play.google.com/store/apps/details?id=es.unizar.gtc.abad.androidlsargv2> (Accedido 18-05-2015).

Implementación de una solución VDI en laboratorios docentes universitarios

Xacobe Macía da Silva, Víctor Manuel Carneiro Díaz, Fidel Cacheda Seijo,
Francisco Javier Nóvoa de Manuel, Diego Fernández Iglesias
Departamento de Tecnologías de la Información y las Comunicaciones
Universidade da Coruña

Facultad de Informática, Campus de Elviña S/N - C.P. 15071 A Coruña

xacobe.macia@udc.es, victor.carneiro@udc.es, fidel.cacheda@udc.es, fjnovo@udc.es, diego.fernandez@udc.es

Resumen—Algunos de los principales problemas de los entornos tradicionales de los laboratorios docentes son los grandes ciclos de puesta en marcha, el elevado número y la complejidad de configuraciones necesarias para las distintas materias y la necesidad de destinar gran número de recursos a estas tareas.

Es en el marco de la búsqueda de soluciones más sencillas, autónomas, eficientes y seguras donde se enmarcan las infraestructuras de escritorios virtuales (“Virtual Desktop Infrastructure” o VDI), que hace referencia al uso de “equipos de escritorio virtualizados” alojados de forma centralizada y aislada en un servidor. De esta forma, el equipo de usuario consiste en una imagen en la estación de trabajo, mientras que los archivos, datos y aplicaciones se almacenan y administran desde un servidor central. Mediante el empleo de *Thin Clients*, dispositivos móviles o equipos comunes, se impulsa claramente la tendencia actual *Bring Your Own Device* (BYOD).

Con el amparo de la experiencia previa de la Universidade da Coruña en la puesta en marcha de una infraestructura VDI para escritorios Windows, en este trabajo, se busca dar respuesta a una serie de necesidades técnicas requeridas por el alumnado y el profesorado de ciertas asignaturas de la Facultad de Informática, como es el caso de la asignatura de Administración de Redes. Dichas necesidades se cubrirán a través de la implantación de una infraestructura de virtualización de escritorios Linux.

El núcleo de este trabajo será, por tanto, realizar un breve análisis de las posibles soluciones VDI para Linux en el mercado actual, seleccionar la adecuada para el entorno docente de la asignatura y realizar su posterior diseño, implantación y despliegue.

Palabras Clave—virtualización, escritorios, linux, vdi, qvd, open source

I. INTRODUCCIÓN

La tecnología VDI se ajusta a la idea de sustituir el equipo de escritorio o PC dedicado en las actividades cotidianas, especialmente en el mundo docente y educativo, donde cada usuario puede emplear su propio dispositivo móvil para la conexión a las redes y servicios de campus. Por esto, es necesario potenciar la presencia de esta tecnología en la Universidade da Coruña, en la que actualmente ya se cuenta con una plataforma de este tipo para equipos Windows.

Así, con el objetivo de dar respuesta a las necesidades docentes, de movilidad del alumnado y el personal docente y buscando un alto rendimiento y optimización de los recursos actualmente disponibles, nace la idea de implantar una solución basada en escritorios virtuales Linux para facilitar la docencia de la asignatura de Administración de Redes, que proporcione la seguridad imprescindible en un entorno de este tipo y que permita dar servicio a los innumerables dispositivos móviles de los que se dispone en la actualidad. De esta forma, se permite mejorar la eficiencia de los equipos Linux necesarios

para impartir la docencia, así como también crear entornos aislados, controlados y personalizables, algo indispensable en una asignatura relacionada con redes y seguridad.

II. ENFOQUE

La asignatura de Administración de Redes es una materia impartida en el segundo cuatrimestre del tercer curso, propia del itinerario de Tecnologías de la Información del Grado en Ingeniería Informática de la Universidade da Coruña, que consta de seis créditos ECTS.

En esta asignatura se asientan los conocimientos sobre redes de ordenadores obtenidos previamente en cursos anteriores, profundizando en el diseño, operación y mantenimiento de redes informáticas y enfocándose a los servicios básicos de red, como la monitorización, la alta disponibilidad, las redes privadas, la gestión de redes y, por supuesto, la seguridad.

Se presenta, por tanto, el objetivo de explicar las necesidades propias de esta asignatura para la realización de las prácticas y justificar la implantación de una infraestructura VDI en este entorno docente; presentar sus ventajas y por supuesto, sus inconvenientes.

II-A. Problemática

La idea de la implantación de una solución VDI surge a partir de ciertas necesidades docentes y administrativas de la asignatura de Administración de Redes, donde se presentan los siguientes requisitos indispensables:

- Necesidad de equipos Linux para la puesta en práctica de las técnicas estudiadas en las clases expositivas, a través de herramientas propias de estos sistemas.
- Necesidad de un entorno de red seguro y aislado del resto de la red de campus, para poner en práctica la configuración y administración de redes, así como también técnicas de seguridad informática.
- Necesidad de gestión y control de todos los alumnos de la asignatura por parte del personal docente.
- Necesidad de disponer de una infraestructura que permita un rápido despliegue y una fácil configuración y gestión de máquinas virtuales, ya que es necesario que cada alumno disponga independientemente de un escritorio o máquina virtual que permita la total manipulación y personalización de la misma.

- Permitir a los alumnos la flexibilidad de realizar las prácticas desde los equipos disponibles en el laboratorio, desde su propio equipo o incluso desde su dispositivo móvil.
- Evitar la necesidad de realizar las prácticas físicamente en el laboratorio, o a través de VPN's, permitiendo que se realicen desde cualquier dispositivo ("Bring Your Own Device") y en cualquier lugar.

II-B. Situación inicial

En un primer momento, las prácticas de la asignatura de Administración de Redes se desarrollaban sobre equipos físicos alojados en un laboratorio y conectados a través de una red LAN aislada del exterior.

Una solución de este tipo presenta una gran variedad de inconvenientes, tanto para el alumnado como para el personal de administración de sistemas. Esta arquitectura obliga al alumnado a llevar a cabo las prácticas exclusivamente en los equipos físicos, únicamente disponiendo de un horario estipulado para el acceso al laboratorio, el cual deberá ser dedicado solamente a esta actividad.

Debido a la naturaleza de las prácticas, la labor de administración era ardua y compleja a consecuencia de las constantes necesidades de restauración de las máquinas a su estado inicial y a la administración manual de los equipos, o en el mejor de los casos, a través de técnicas de clonación.

II-C. Punto de partida

La situación inicial anterior, poco a poco fue sustituida por una más acorde a las necesidades, que facilitase la gestión de los equipos y mejorase la flexibilidad de los usuarios. La situación anterior a la implantación del proyecto VDI era la siguiente:

- Creación de una máquina virtual independiente por alumno matriculado, a través de la virtualización de servidores (*VMware vSphere*). Evidentemente, la creación y gestión de las máquinas virtuales eran tareas llevadas a cabo por el administrador de la infraestructura, de forma manual.
- Administración y configuración de todo el hardware de forma independiente.
- Indisponibilidad hardware para la ejecución de la totalidad de máquinas virtuales (aproximadamente doscientas máquinas) de forma simultánea, de tal forma que era necesario parar y arrancar las máquinas de forma manual en función del horario de la asignatura, tarea realizada por el personal docente. Para realizar una práctica fuera de horario de la asignatura, los profesores debían arrancar las máquinas a petición de los alumnos.

Dicha situación supuso el punto de partida previo a la implantación del proyecto de virtualización de escritorios.

III. OBJETIVOS

El objeto de este proyecto es analizar y posteriormente implantar una infraestructura de virtualización de escritorios Linux, tratando también de explicar de forma breve y concisa el amplio panorama de la temática VDI en general.

Consistirá básicamente en el diseño de la infraestructura, y su posterior configuración y despliegue de una plataforma plenamente funcional, tras pasar por la perspectiva de un proyecto piloto para evaluación por parte del equipo docente encargado de impartir la asignatura de Administración de Redes.

Además, se pretende cumplir los siguientes objetivos, complementarios a lo anterior:

- Realizar un análisis y una posterior selección entre varias plataformas de virtualización de escritorios Linux existentes para poder escoger la que mejor se adapte a los requisitos de la Universidade da Coruña.
- Realización del proyecto teniendo siempre presentes las posibles necesidades futuras: implantar y diseñar una infraestructura eficiente, altamente disponible y fácilmente escalable.

IV. DESARROLLO

IV-A. Alternativas VDI

Son muchas las soluciones existentes en el mercado actual para llevar a cabo la implantación de un sistema de virtualización de escritorios Linux. Es en este momento donde se presenta una de las cuestiones más importantes del proyecto: elegir la adecuada para la asignatura.

Para llevar a cabo esta decisión, se contó con el apoyo del Servicio de Informática y Comunicaciones de la Universidade da Coruña, cuya experiencia previa en la implantación de la solución de *VMware (VMware Horizon View)* permitió realizar un primer filtrado para seleccionar tres alternativas susceptibles de ser analizadas en detalle: *Ulteo* [1], *QVD* [2] y *Red Hat (Red Hat Virtualization for Desktops)* [3]). A continuación se describirá cada una de las alternativas:

- **Ulteo:** Se trata de una solución muy completa que ofrece clientes de acceso a los escritorios virtuales para todas las plataformas, incluso un cliente de acceso web HTML5 y Java. Destaca también por poseer una interfaz de administración muy completa y con gran cantidad de funcionalidades.

Además, permite la vinculación con sistemas de autenticación externa como LDAP o *Active Directory*, algo indispensable en un entorno con un elevado número de usuarios.

Una de las características más importantes de la solución es la disponibilidad de aplicaciones SaaS¹ o integración de aplicaciones, que consiste en la ejecución de aplicaciones del escritorio virtual desde el escritorio local, como si de una aplicación nativa se tratase. Además ofrece algunas otras interesantes funcionalidades como "*Dynamic delivery fallback*", que permite la instalación de aplicaciones de forma dinámica en los escritorios virtuales. Por ejemplo, si se necesita instalar una determinada aplicación en todos los escritorios, esto se realiza de forma automática y sin interacción por parte del usuario. Es importante remarcar que estas características de Ulteo no se encuentran disponibles en todas las versiones.

En sus versiones de suscripción², Ulteo incluye un soporte

¹"Software as a Service", modelo de distribución de software donde el soporte lógico y los datos que maneja se alojan en servidores de una compañía de tecnologías de información y comunicación (TIC).

²Se recuerda que también son analizadas por las posibles necesidades de ampliación futura del proyecto.

de nivel tres³, es decir, un soporte de disponibilidad completa y personalizada las veinticuatro horas del día, los siete días de la semana y durante todo el año (24x7x365).

- **QVD:** Es la segunda de las alternativas analizadas. QVD es un proyecto "Open Source". Es una solución que ofrece clientes de acceso para todas las plataformas (excepto iOS, que se encuentra en desarrollo). Es una alternativa que ofrece exclusivamente virtualización de sistemas Linux, a diferencia de *Ulteo*, que permite también Windows. Algo que aventaja a *QVD* por encima de las otras soluciones es la disponibilidad de la totalidad de funcionalidades en su versión gratuita. Cabe destacar también que ofrece una interfaz web más sencilla, pero menos completa que la de *Ulteo* (en apartados posteriores se realizará una comparativa de forma específica). El sistema de soporte de *QVD* difiere levemente con *Ulteo*, ya que dispone de dos planes de soportes diferenciados en función de la versión seleccionada: soporte siguiente día laboral (5x8) o soporte completo (24x7) en su versión más completa.
- **Red Hat Virtualization for Desktops:** Es la solución VDI ofrecida por la conocida empresa Red Hat Inc. Destaca por el uso de un protocolo de comunicación escritorio-servidor propio, SPICE [5]. Se trata de la menos completa de las soluciones analizadas, en lo que a variedad de clientes de acceso se refiere. Esta solución, destaca por hacer uso a bajo nivel de los servidores de virtualización *Red Hat*, permitiendo así complementarse a la perfección con infraestructuras que ya hagan uso de los mismos y otorgando un gran rendimiento, así certificado en las pruebas *SPECvirt*⁴ [6], incluso con un elevado número de máquinas virtuales en un nodo servidor. Por otra parte, permite la virtualización tanto de escritorios Windows como Linux y, al igual que *Ulteo*, ofrece la posibilidad de entregar aplicaciones SaaS. Del mismo modo que *QVD*, *Red Hat* permite la contratación de un soporte *business-hour*⁵ (8x5) o por el contrario un soporte completo (24x7).

IV-B. Selección

Una vez elaborado el análisis de cada una de las posibles alternativas, es necesario realizar una exhaustiva comparativa de las mismas, poniendo especial atención en aquellas características indispensables, por ser necesarias en el entorno en el que se implantará la infraestructura. Asimismo, se omiten de la comparativa las características innatas de los sistemas analizados o aquellas que pueden darse por supuestas (p.ej: alta densidad, alta disponibilidad, balanceo de carga o compatibilidad con

³Un soporte de nivel tres es el que ofrece la mayor capacidad de resolución de problemas. Estipula que cualquier problema será atendido por profesionales expertos y dedicados exclusivamente, así como también personal de investigación y desarrollo para crear soluciones nuevas.

⁴Se trata de una suite de benchmarks para medir el rendimiento de plataformas de virtualización, desarrollados por un conglomerado de empresas hardware, software e investigadores.

⁵Plan de soporte durante horario laboral.

sistemas de autenticación externa, son características que todas las soluciones poseen).

Dadas las características del entorno en el se realizará la instalación del sistema, la solución ideal sería la implantación de cualquiera de las soluciones en sus versiones "enterprise" destinadas a organizaciones con grandes infraestructuras informáticas y que precisan un soporte de calidad. Por ser este un proyecto piloto para evaluación por parte del personal docente y el alumnado, la implantación del sistema será realizada, en una primera instancia, a través de alguna de las versiones gratuitas de las soluciones analizadas.

Teniendo en cuenta una posible ampliación y puesta en producción en un futuro próximo, la comparativa real se realizó considerando también las versiones completas. De este modo se puede tomar una decisión final objetiva y siempre considerando la posibilidad de que en un futuro la infraestructura sea migrada hacia alguna de las versiones completas y con soporte.

IV-C. Solución a implantar

Tras la realización de la comparativa, y debido al condicionante inicial de implantación de la infraestructura sin partida presupuestaria, la alternativa de *Red Hat* debió ser descartada completamente por dos motivos fundamentales: el primero de ellos es la obligatoriedad por parte de *Red Hat* de realizar la implantación en servidores *Red Hat Enterprise Linux*, por lo que es necesario disponer de servidores con dicha licencia. El segundo de los motivos es la no disponibilidad de una alternativa gratuita de la solución *Red Hat Virtualization for Desktops*, ya que sólo dispone de una evaluación de sesenta días, insuficiente para la puesta en producción y evaluación por parte del personal docente de la asignatura durante un periodo lectivo. Es por estos motivos por los que únicamente se realizó el estudio práctico de las soluciones de *QVD* y *Ulteo*.

Las pruebas iniciales realizadas consistieron en distintos test con las máquinas virtuales o "appliances" preconfiguradas que ambas soluciones proporcionan para dicho fin. Los análisis fueron los siguientes:

- **Análisis de la interfaz web de administración:** Este estudio consistió en comparar las diferentes opciones y herramientas de administración que incluye la interfaz web. Dentro de este campo las características más importantes que se tuvieron en cuenta fueron: sencillez, completitud y usabilidad. En este apartado la interfaz ofrecida por *Ulteo* vence a *QVD* pues desde la primera de ellas se puede realizar la totalidad de las tareas de administración, mientras que en la solución *QVD*, la administración debe ser complementada con la línea de comandos. Inicialmente, esto no supone un inconveniente si el nivel técnico de los administradores es alto.
- **Análisis de los clientes software:** Para la realización de esta prueba se instalaron cada uno de los clientes disponibles en todas sus plataformas:
 - **Ulteo:** La solución *Ulteo* en su versión gratuita ofrece clientes de acceso al escritorio virtual a través del navegador, bien mediante el motor Java, como recientemente a través de HTML5. Los clientes nativos y los clientes para plataformas móviles sólo están disponibles en alguna de sus versiones de pago.

- **QVD:** A diferencia de Ulteo, en este caso resulta mucho más completa la solución QVD, pues permite la instalación y uso de todos los clientes disponibles, excepto el cliente iOS, que se encuentra en desarrollo.

▪ **Análisis de rendimiento de los escritorios:** Se trata de una prueba de difícil justificación inicial, ya que en el uso de las máquinas virtuales servidor que las soluciones ofrecen para pruebas, ambas soluciones ofrecen un rendimiento similar. El problema reside en la imposibilidad de realizar una prueba de ambas soluciones en un entorno real, con una carga de trabajo real (varias decenas de usuarios). Es por ello que en este apartado se considerará el rendimiento de ambas soluciones equivalente.

▪ **Análisis de otras características:** En el resto de las características de las versiones gratuitas, prácticamente no se aprecian diferencias significativas, pues ambas soluciones cubren completamente el campo de elementos necesarios para una solución VDI básica: balanceo de carga, alta disponibilidad, control de acceso (“Active Directory” o LDAP), autoprovisión de escritorios, gestión centralizada, etc.

Aunque desde un primer momento Ulteo se posiciona de una forma aventajada, la carencia de clientes de acceso nativos en su versión gratuita hace que no sea tan adecuada como QVD para este proyecto, pues esta última permite y flexibiliza a los alumnos y profesorado el uso de escritorios virtuales desde cualquier lugar, algo primordial y requisito imprescindible.

Bien es cierto que QVD ofrece una interfaz de administración web menos completa y con menos funcionalidades (únicamente permite la realización de tareas básicas como crear/eliminar usuarios, máquinas virtuales e imágenes de disco), pero que se complementa con una serie de *scripts* de fácil uso para el resto de tareas. Por todo esto, finalmente se consideró QVD como la solución ideal para este proyecto, para favorecer a los usuarios el acceso a los escritorios virtuales, en detrimento de una administración más compleja por parte de los administradores.

IV-D. QVD

QVD es la solución de escritorios virtuales ofrecida por la empresa *Qindel Group*, una compañía internacional de consultoría especializada en proyectos y entornos Linux.

El software está diseñado íntegramente para ofrecer virtualización de escritorios Linux. Los clientes se conectan a un servidor o servidores centrales que son los encargados de cargar el entorno y las aplicaciones propias de cada usuario.

En lo referente a las tecnologías empleadas por QVD, destaca por ser una de las únicas soluciones VDI para Linux que ofrece la posibilidad de seleccionar la tecnología de virtualización a bajo nivel deseada: KVM [7] o LXC [8], permitiendo implementar cualquiera de las dos opciones, en función de las necesidades de la organización o la preferencia de los administradores.

Los sistemas operativos son cargados a través de imágenes de disco independientes y totalmente personalizables en los nodos de la infraestructura QVD. De esta forma se puede crear un entorno personalizado para un usuario o un grupo de usuarios,

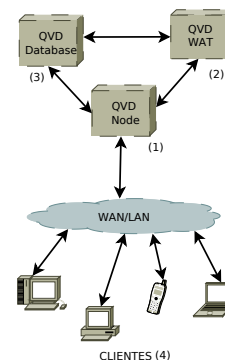


Figura 1. Esquema básico de QVD

algo que se adapta a la perfección para el entorno docente al que se orienta este proyecto.

El acceso al escritorio virtual se realiza a través de un cliente de acceso, previa autenticación (local o centralizada) y mediante el uso del protocolo NX [10], añadiendo así una capa de seguridad adicional al proceso al viajar el tráfico cifrado a través de SSH, pudiendo acceder de forma segura y desde cualquier dispositivo.

IV-D1. Componentes principales: Para construir una solución completa de escritorios virtuales de QVD es necesaria la presencia de una serie de componentes indispensables para el correcto funcionamiento de la plataforma, tal y como se puede ver en la Figura 1. Dichos componentes pueden estar presentes de forma conjunta en un solo servidor físico, o bien distribuirse en diferentes nodos. Evidentemente, las instalaciones monoservidor están destinadas únicamente a pequeños entornos o para la realización de pruebas. Los componentes principales son los siguientes:

- **Servidor QVD (1):** Se trata del nodo donde se ejecutan las máquinas virtuales o escritorios virtuales. Normalmente serán una serie de servidores formando una arquitectura *clúster*.
- **Servidor de administración (2):** Es el encargado de ejecutar el servidor en el que se ejecuta la interfaz de administración web, que emplean los administradores para realizar las tareas de gestión.
- **Servidor de base de datos (3):** Se trata de un servidor *PostgreSQL*, que contiene una base de datos donde se almacena toda la información acerca de los usuarios, máquinas, equipos, sistemas operativos y configuraciones del entorno.
- **Clientes de acceso (4):** Se trata del conjunto de programas software disponibles para las diferentes plataformas y que permiten el acceso a los escritorios virtuales.

IV-D2. Componentes secundarios: Además de los componentes indispensables citados en la sección anterior, existen otros que pese a no ser completamente necesarios, están igualmente presentes en la práctica totalidad de

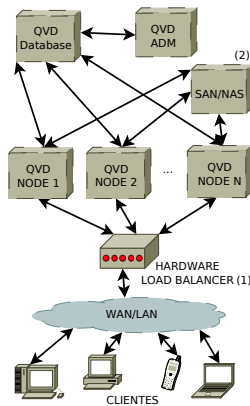


Figura 2. Componentes adicionales de QVD.

las infraestructuras con un mínimo de usuarios. Dichos componentes se muestran en la Figura 2.

- **Balancedor de carga (1):** Necesario en cualquier entorno en producción en el que se ejecutan múltiples máquinas virtuales para dar servicio a varios usuarios. Dicho balancerador será el encargado de redistribuir equitativamente las máquinas virtuales en ejecución entre los nodos QVD.
- **Almacenamiento compartido (2):** En una infraestructura en la que se dispone de más de un nodo QVD, todos los componentes deben tener acceso a una zona de almacenamiento compartido (normalmente NFS) en el que se almacenan imágenes de los sistemas operativos y los datos de los usuarios para cargar las máquinas virtuales.
- **QVD Virtual Machine Agent:** Es el responsable de aceptar las conexiones desde el cliente a la VM a través del nodo QVD. Técnicamente, es un demonio Linux que se instala en cada una de las máquinas virtuales y es el encargado de permitir la comunicación entre los nodos físicos y las máquinas virtuales que está ejecutando el nodo.

IV-D3. Componentes internos: Profundizando en la arquitectura de QVD, es el momento de tratar los componentes que lo conforman a bajo nivel, los realmente encargados de llegar a cabo las tareas principales de la infraestructura.

En la Figura 3 se muestra un diagrama completo en el que se introducen los nuevos componentes y las relaciones de los mismos con los componentes descritos en las secciones anteriores. A continuación se describe cada uno de ellos:

- **L7R:** Conocido como *Layer-7 Router*. Es el encargado de ejercer las tareas de *broker*⁶, característico de este tipo de entornos VDI. Es el responsable de realizar dentro de los nodos QVD todas las tareas para establecer y gestionar las sesiones que los usuarios abren contra sus máquinas virtuales. **Gestiona los usuarios.**

⁶Encargado de determinar qué escritorio virtual será asignado a qué usuario, siendo éste el que establece la conexión.

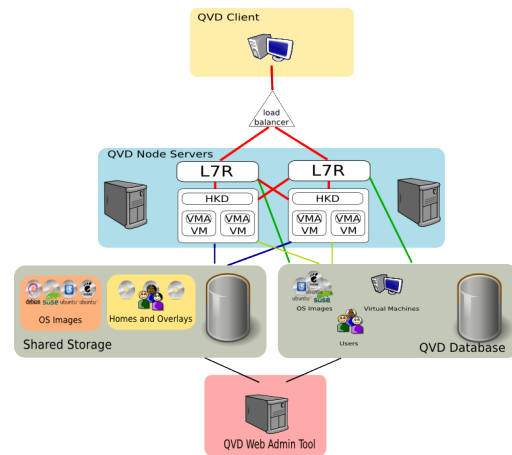


Figura 3. Arquitectura general de QVD.

- **HKD:** Es el otro de los demonios Linux encargados del funcionamiento interno de QVD. El *House Keeping Daemon* es el encargado de gestionar el estado de las máquinas virtuales (arranques, paradas, etc). Gestiona y monitoriza cada una de las VM, siendo también el responsable de actualizar la información sobre el estado de cada máquina en la base de datos. **Gestiona las máquinas virtuales.**

IV-E. Diseño del sistema

Una vez conocida la arquitectura típica y componentes de la solución de QVD, es el momento de llevar a cabo la toma de otras decisiones que condicionarán la instalación y diseño de la infraestructura.

Las decisiones importantes en este punto son: el entorno y la arquitectura de QVD para el proyecto, la tecnología de virtualización a emplear y el sistema operativo seleccionado para los nodos.

IV-E1. Selección de la tecnología: Analizada en detalle la documentación de QVD, el primero de los pasos, previo a la instalación es la decisión de la tecnología de virtualización a emplear: KVM [7] o LXC [8]. Es uno de los pasos más importantes en el proceso de implantación de QVD y que condicionará todo el proceso y el tratamiento de las imágenes de disco de las máquinas virtuales.

El punto clave para la toma de decisión, radica en la compatibilidad del kernel de los sistemas operativos. Los sistemas operativos virtualizados sobre LXC deben ser compatibles con el kernel del host anfitrión, es decir, los sistemas operativos de las máquinas virtuales usan exactamente el mismo kernel que la máquina anfitriona (host físico) tal y como se describe en el manual técnico de la tecnología [8]. Esto hace que sea altamente recomendable el uso de la misma distribución en las máquinas virtuales y en el equipo físico, algo que limita la flexibilidad de selección de operativo por parte de los usuarios (todos los escritorios virtuales deben correr el mismo operativo que los servidores físicos). Esto no sucede en la tecnología KVM, donde el kernel del equipo físico y el de las máquinas virtuales son completamente independientes [7].

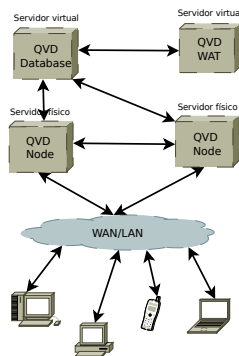


Figura 4. Arquitectura de QVD implementada

A consecuencia de la compartición de kernel que realiza LXC, éste ofrece un rendimiento cercano al nativo en la ejecución de las máquinas virtuales, mientras que el uso de kernel independiente de KVM, provoca que sea necesaria una comunicación y traducción de instrucciones entre ambos, degradando así el rendimiento. De este modo, en el mismo hardware, es posible ejecutar mayor número de instancias mediante el uso de la tecnología LXC.

Son, por tanto, la poca flexibilidad que permite LXC y la menor información existente sobre la tecnología los puntos que declinaron la balanza hacia KVM como tecnología de virtualización a emplear en el proyecto.

IV-E2. Diseño de la arquitectura: Para la realización del diseño físico de la infraestructura QVD, existen una serie de condicionantes que marcan el camino a seguir: el presupuesto, la disponibilidad de equipos hardware y la posibilidad de ampliación futura.

Al ser este un proyecto inicial y piloto, no se realiza la compra de equipos nuevos completamente dedicados a la infraestructura, sino que se hará uso de equipos físicos en desuso y de servidores virtuales.

En la Figura 4 se muestra el diseño implementado para este proyecto, que se describe a continuación:

- **Un servidor virtual para la base de datos:** En una infraestructura QVD, la BD constituye un componente indispensable, ya que almacena y mantiene todos los datos necesarios para el vital funcionamiento del conjunto. A pesar de esto, su carga de trabajo no es excesivamente elevada (sólo se producen ciertas consultas en el proceso de inicio y actualizaciones periódicas del estado de las VM), por lo que un servidor virtual con unos requisitos moderados será suficiente para albergarla.

Como es lógico, en función de la carga de usuarios se realizará el dimensionamiento del servidor, aunque inicialmente se ha creado un servidor virtual⁷ con un

⁷La creación de servidores virtuales se lleva a cabo mediante la infraestructura de virtualización de servidores VMware vSphere 5.5.

dimensionamiento mínimo⁸, pero al tratarse de un servidor virtual, es altamente escalable en el momento que las necesidades así lo requieran⁹.

- **Un servidor de administración virtualizado:** Suficiente para la ejecución de un servidor Apache para la aplicación web de administración. Debe ser capaz de soportar un número reducido de conexiones a la aplicación web y a través de SSH, que realizarán los administradores para el mantenimiento y monitorización de la infraestructura. Al igual que en el caso anterior, se trata de un servidor virtual con un dimensionamiento moderado (una CPU de un núcleo y 2GB de RAM), pero escalable en función de las necesidades.
- **Dos servidores físicos para los nodos QVD:** Los servidores físicos, de mayor potencia hardware, serán los encargados de alojar los nodos QVD, sobre los que se ejecutarán los escritorios virtuales. La escalabilidad de la infraestructura se ve penalizada en este punto por los equipos hardware disponibles. Para una primera instalación inicial, destinada a pruebas, se emplearon dos servidores físicos en desuso¹⁰, aunque una vez realizada la completa instalación, se migraron hacia equipamiento más potente¹¹.

IV-E3. Diseño de la red: Diseñada la arquitectura de los sistemas, es el momento de diseñar las redes de comunicación necesarias en la infraestructura, en el que la única labor radica en el diseño y los requisitos necesarios, ya que la tarea de creación y administración de la misma pertenece al Departamento de Red y Comunicaciones.

- **Red de administración:** Se trata de la red necesaria para albergar y comunicar los nodos físicos, el servidor de BD y el servidor de administración.
- **Red de máquinas virtuales:** Es la red en la que se encontrarán todas las máquinas o escritorios virtuales que se creen en la infraestructura. Está vinculada a la red de administración a través de un bridge en cada uno de los nodos físicos. Es importante destacar que esta red debe disponer de direcciones IP suficientes para albergar a cada uno de los escritorios virtuales.

IV-F. Características adicionales

Una vez realizada la instalación básica de la infraestructura, y realizadas las pruebas de funcionamiento, se llevó a cabo la introducción de una serie de mejoras para la optimización del conjunto.

- **Autenticación externa:** Aunque QVD proporciona su propio framework de autenticación, el uso de un mecanismo

⁸Una CPU de un núcleo y 1GB de RAM.

⁹QVD en su manual técnico establece que el uso de un servidor con 2 núcleos CPU y 2GB de RAM es suficiente para soportar la carga de cualquier entorno de producción.

¹⁰HP ProLiant DL140 G3. 2 x Intel(R) Xeon(R) CPU 5110 1.60GHz 64-bit. Memoria 3GB RAM.

¹¹HP ProLiant DL785 G5. 16 x Quad-Core AMD Opteron(tm) CPU 8354 2.2GHz 64-bit. Memoria 128GB RAM.

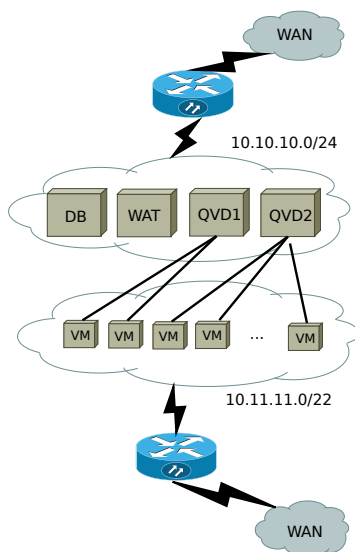


Figura 5. Arquitectura de red QVD.

de autenticación externo como LDAP es algo común en organizaciones de cierto tamaño.

El servicio LDAP corporativo de la Universidad, es el lugar donde se mantienen centralizados y clasificados una serie de datos de los miembros de la comunidad universitaria. De esta forma, una vez configurado el mecanismo en QVD, cualquier usuario matriculado en la asignatura correspondiente puede autenticarse en la infraestructura.

- **Autoprovisión:** Por sí solo, el mecanismo de autenticación externa no facilita el trabajo de los administradores, ya que aunque no sea necesario crear los usuarios manualmente, estos sí deben crear y asignar de forma manual una máquina virtual a cada usuario. Es por este motivo que la autenticación externa debe ser combinada con la autoprovisión, que consiste en asignar automáticamente una máquina virtual a un usuario. Es decir, cuando un alumno se autentica por primera vez en la infraestructura, el sistema de autoprovisión genera una nueva máquina virtual y la asigna al alumno de forma transparente. Evidentemente, este proceso puede ser configurado y personalizado por los administradores (características de la máquina virtual, imagen de disco a utilizar, recursos hardware, etc).
- **Balancedor de carga:** Se trata de una característica necesaria en un entorno VDI de alta capacidad. Sin configuraciones adicionales, QVD integra a través de software, un balancedor de carga interno, implementado en el demonio L7R, que consiste en la asignación de diferentes valores o pesos a características (RAM, CPU, etc) y en

función de la carga de trabajo instantánea en cada uno de los nodos, el L7R ejecutará la VM en un nodo u otro.

Además del propio balancedor de carga interno, se llevó a cabo la instalación de otro balancedor externo para desempeñar la tarea de *broker*, característico de cualquier solución VDI. Básicamente es un equipo independiente encargado de aceptar las peticiones de los clientes. De esta forma, todos los clientes realizarán sus conexiones contra una misma dirección IP¹² y será este balancedor o *broker* el encargado de enviar la petición del cliente a uno de los nodos a través de una asignación *Round-Robin*¹³. Dicho *broker*, previamente a realizar el envío de la petición del cliente, realiza una comprobación de vida de los nodos, a través de un mecanismo implementado en QVD conocido como *“health checking”*. Este mecanismo consiste en enviar una petición HTTP GET al nodo el cual responde con un mensaje de confirmación.

- **Configuración SSL:** Para aumentar la seguridad de las conexiones se ha configurado el empleo de HTTPS para el proceso de conexión con la infraestructura, a través de un certificado x509 y clave privada.
- **Alta disponibilidad:** La instalación inicial de los nodos QVD se realizó en dos equipos hardware situados en la misma localización física, pero tras las primeras pruebas de la infraestructura, se procedió a migrar los nodos a otras máquinas físicas más potentes, anteriormente mencionadas. Con estos dos nuevos servidores físicos, además de mejorar la potencia y la capacidad de dar servicio a un mayor número de usuarios (como veremos en la prueba mostrada en la siguiente sección), se mejora también la disponibilidad del entorno, al formar una arquitectura altamente disponible. Entre la localización física de los dos servidores existen ahora varios kilómetros de distancia.

IV-G. Pruebas de carga

Al comenzar a trabajar con infraestructuras VDI, son muy frecuentes las sorpresas en lo que a degradación de rendimiento se refiere; es decir, no obtener el rendimiento esperado inicialmente por parte de los servidores hardware. Esto es producido por las grandes demandas de rendimiento del almacenamiento que exigen las soluciones VDI, por lo que se hace indispensable un buen diseño y una gran inversión en discos o cabinas de discos.

Estos problemas relacionados con la E/S son conocidos como tormentas de IOPS¹⁴ y básicamente consiste en cargas elevadas de procesos de E/S en momentos puntuales, como el inicio de sesión de decenas de equipos de forma simultánea.

Como es evidente, uno de los momentos con mayor exigencia al almacenamiento es el proceso de arranque de las máquinas virtuales. En la Figura 6 se muestra una gráfica de las pruebas realizadas, donde se observa el tiempo de arranque simultáneo

¹²O una dirección DNS, en este caso

¹³Consiste en la selección de elementos empezando por el primero de la lista hasta llegar al último. Este proceso se repite de manera iterativa.

¹⁴Input/Output Operations Per Second: Es la unidad benchmark empleada para medir el rendimiento de los discos duros o cualquier otra unidad de almacenamiento.

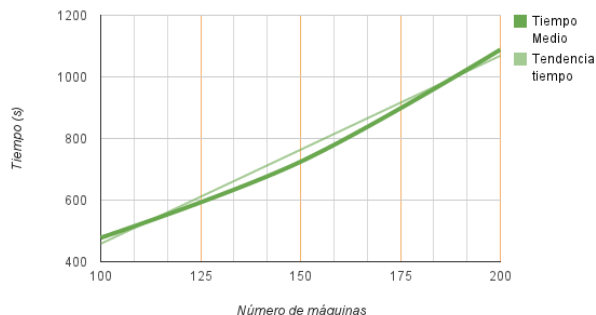


Figura 6. Relación de tiempo de arranque y número de máquinas virtuales.

de todas las VM de la infraestructura variando el número de las mismas. Tal y como se puede apreciar, se realizaron distintas pruebas con un número diferente de máquinas entre las cien y las doscientas, obteniendo resultados satisfactorios y perfectamente asumibles (se logra arrancar doscientas máquinas en menos de diecinueve minutos) a pesar de tratarse de un almacenamiento compartido vía NFS, a través de un servidor con reducidos recursos hardware¹⁵, suficiente para este proyecto de evaluación de la solución¹⁶.

V. TRABAJO FUTURO

Tras la finalización de este proyecto, se abren una serie de líneas de continuidad que permiten establecer objetivos a largo plazo:

- **Automatización de tareas:** Elaboración de *scripts* que permitan realizar de forma automática tareas de administración características de un entorno de este estilo, como por ejemplo: permitir eliminar, realizando previamente una copia de seguridad, todas las máquinas virtuales al finalizar el curso académico y dejar la infraestructura perfectamente preparada para el inicio del próximo período lectivo.
- **Establecer una política de backup:** Aunque durante el período de puesta en marcha y actualmente durante la evaluación se realizan copias de seguridad de forma manual, es importante establecer una estricta política de seguridad que permita clasificar las copias de las máquinas virtuales por curso académico (derivado del punto anterior) y permita devolver la infraestructura al estado deseado ante cualquier problema o desastre.
- **Mejora de las imágenes de disco:** Creación y mejora de nuevas imágenes de disco personalizadas, con software completamente adaptado a las necesidades del personal docente que imparte la asignatura de Administración de Redes.
- **Búsqueda del Thin Client ideal:** Realizar un exhaustivo análisis de las posibles alternativas de equipos ligeros para dar con la solución ideal (coste-rendimiento) para sustituir, por ejemplo, los equipos convencionales de los laboratorios

¹⁵Un servidor virtual VMware vSphere con un núcleo CPU y 2GB de RAM.

¹⁶Evidentemente, una vez se realice la puesta en producción definitiva de la infraestructura, se debe tener muy en consideración la compra y diseño del almacenamiento.

de prácticas por equipos *Thin Client* y abaratar así el coste de los equipos y de mantenimiento.

- **Extender el uso de la solución:** Evaluar y analizar la posibilidad de ampliar esta infraestructura para otras asignaturas o para cualquier otro campo del entorno universitario donde sea necesario el uso de escritorios Linux, como la investigación.

VI. CONCLUSIONES

En este trabajo se ha presentado una nueva infraestructura para facilitar y dar servicio a los alumnos de la asignatura de Administración de Redes, construyendo un completo y aislado entorno virtual en el cual pueden realizar las prácticas de la asignatura con total libertad y flexibilidad: libertad horaria y flexibilidad en cuanto a dispositivos y localización física.

La acogida de la solución tanto por parte del profesorado como del estudiantado, está siendo inicialmente muy positiva, a pesar de ser recientemente implantada. A falta de una valoración clara y concisa tras la finalización del curso, las primeras impresiones muestran una clara adaptabilidad técnica e ideológica del proyecto, pues ofrece la perfecta simulación del entorno real, reduciendo los costes y facilitando las tareas del profesorado en cuanto a administración del laboratorio se refiere.

La gran acogida de esta solución permite ampliar el horizonte hacia la posible implantación de una infraestructura VDI común para todos los laboratorios de la Facultad de Informática, con grupos de equipos completamente personalizados y destinados a cada una de las asignaturas que son impartidas.

AGRADECIMIENTOS

La realización de este trabajo fue posible gracias a la Consejería de Cultura, Educación e Ordenación Universitaria con su ayuda GPC2014/045 (Grupos con Potencial Crecimiento).

Los autores quieren agradecer también la aportación de los profesores del Grupo de Telemática de la Universidad de Coruña, así como también la participación y experimentación en el proyecto del personal de la asignatura de Administración de Redes.

REFERENCIAS

- [1] AZNetwork SAS, *Ulteo* [En línea]. France: AZNetwork SAS, 2015 [Consulta: 10 de Marzo de 2015]. Disponible en: <http://www.ulteo.com/home/en/home>
- [2] Qindell Group, *The QVD* [En línea]. Madrid: Qindell Group, 2015 [Consulta: 17 de Marzo de 2015]. Disponible en: <http://theqvd.com/es/>
- [3] Red Hat Inc. , *Red Hat Virtualization for Desktops* [En línea]. USA: Red Hat Inc. , 2015 [Consulta: 16 de Marzo de 2015]. Disponible en: <http://www.redhat.com/es/technologies/virtualization>
- [4] VMware Inc. , *VMware Horizon View* [En línea]. USA: VMware Inc., 2015 [Consulta: 4 de Marzo de 2015]. Disponible en: <http://www.vmware.com/>
- [5] Red Hat Inc. , *Red Hat SPICE* [En línea]. USA: Red Hat Inc. , 2015 [Consulta: 16 de Marzo de 2015]. Disponible en: <https://www.redhat.com/es/resources/red-hat-enterprise-virtualization-and-spice-protocol>
- [6] Standard Performance Evaluation Corporation *SPECvirt_sc@2013* [En línea]. USA: Standard Performance Evaluation Corporation, 2013 [Consulta: 24 de Marzo de 2015]. Disponible en: <https://www.spec.org/>
- [7] Kernel-based Virtual Machine Project [En línea]. Open source project, 2015 [Consulta: 29 de Marzo de 2015]. Disponible en: http://www.linux-kvm.org/page/Main_Page
- [8] Linux Containers Project [En línea]. Open source project [Consulta: 29 de Marzo de 2015]. Disponible en: <https://linuxcontainers.org/>
- [9] Qindell Group, *The QVD Administration Manual* [En línea]. Madrid: QVD Docs Team, 2015 [Consulta: 17 de Marzo de 2015]. Disponible en: <http://docs.theqvd.com/docs/AdministrationManual.html>
- [10] NoMachine S.r.l. , *NoMachine NX* [En línea]. Francia: NoMachine S.r.l., 2015 [Consulta: 29 de Marzo de 2015]. Disponible en: <https://www.nomachine.com/AR11K00745>

Diseñando un MOOC en edX: Introducción a la Programación con Java – Parte 1

Carlos Alario-Hoyos, Carlos Delgado Kloos, Iria Estévez-Ayres, Carmen Fernández-Panadero, Jorge Blasco, Sergio Pastrana, Guillermo Suárez-Tangil, Julio Villena-Román

Departamento de Ingeniería Telemática y Departamento de Informática
Universidad Carlos III de Madrid

Avda. de la Universidad 30, 28911, Leganés (Madrid), España

calario@it.uc3m.es, cdk@it.uc3m.es, ayres@it.uc3m.es, mcfp@it.uc3m.es, guizos@gmail.com,
spastran@inf.uc3m.es, guillermo.suarez.tangil@gmail.com, jvillena@it.uc3m.es

Resumen- Los MOOCs (*Massive Open Online Courses* – Cursos Online Masivos Abiertos) han supuesto una revolución en la educación. Diseñar y poner en marcha un MOOC es una tarea compleja y además dependiente fuertemente del tipo de curso y de la plataforma donde éste se vaya a desplegar. Este artículo presenta el diseño del MOOC “*Introducción a la Programación con Java – Parte 1: Empezando a Programar en Java*” desplegado en la plataforma edX y puesto en marcha a finales de abril de 2015. Este curso de cinco semanas introduce los principios básicos de la programación utilizando como herramienta el lenguaje Java. El MOOC ha surgido de la colaboración entre los departamentos de Ingeniería Telemática e Informática de la Universidad Carlos III de Madrid. En el momento de su apertura el curso contaba con más de 60.000 alumnos de 190 países (según los datos proporcionados por edX) y en el momento de la escritura de este artículo el MOOC todavía se encontraba en el periodo de impartición. Además de presentar el diseño del curso, este artículo extrae algunas de las principales lecciones aprendidas por parte del equipo docente tras el diseño y puesta en marcha de este MOOC.

Palabras Clave- MOOC, edX, programación, Java.

I. INTRODUCCIÓN

Los MOOCs (*Massive Open Online Courses* – Cursos Online Masivos Abiertos) han supuesto un cambio muy importante en la forma en la que se aprende y se enseña hoy en día en la práctica totalidad de áreas de conocimiento [1]. Los MOOCs han supuesto una oportunidad para acceder a contenidos de alta calidad sobre los que construir nuestro aprendizaje acerca de una determinada materia, y al mismo tiempo han permitido crear un escaparate en el que los profesores pueden extender su docencia fuera del aula, a lo largo de todo el mundo [2]. Además, las tecnologías, formatos y lecciones aprendidas de los MOOCs, pueden utilizarse también para mejorar la docencia en el campus a través de los conocidos como SPOCs (*Small Private Online Courses* – Cursos Online Pequeños Privados) [3].

La Universidad Carlos III de Madrid (UC3M) ha apostado fuertemente por incorporar los MOOCs (y los SPOCs) a su oferta educativa desde 2012 [4][5], participando en dos de las principales iniciativas MOOC: edX [6] y MiríadaX [7]. Las primeras experiencias con MOOCs se llevaron a cabo a principios de 2013 en la plataforma MiríadaX ofreciendo tres cursos en español [8]. Desde febrero de 2014 la UC3M forma parte de edX, en cuya plataforma ha comenzado a lanzar sus primeros cursos en inglés y español a principios de 2015.

El apoyo de la UC3M a la creación de MOOCs (y SPOCs) entre su personal docente se ha materializado también en la creación de la Unidad de Tecnología Educativa e Innovación Docente (UTEID) [9], cuyo objetivo principal es apoyar a profesores que desean poner en marcha MOOCs y SPOCs asesorando en cuestiones tecnológicas, audiovisuales, de creación de contenidos, de reutilización de contenidos de terceros o de despliegue en diferentes plataformas, entre otras. La labor realizada por la UTEID se ha recogido y publicado en forma de buenas prácticas [10][11][12][13].

En este artículo se presenta el diseño del MOOC “*Introducción a la Programación con Java – Parte 1: Empezando a Programar en Java*” (“*Introduction to Programming with Java - Part 1: Starting to Programming in Java*”) [14] realizado a través de la colaboración entre los departamentos de Ingeniería Telemática e Informática de la UC3M y con el apoyo de la UTEID y desplegado en la plataforma edX (sección II). Además, en este artículo se recogen un conjunto de lecciones aprendidas por el equipo docente durante el proceso de diseño y puesta en marcha de este MOOC (sección III). El artículo finaliza recogiendo las conclusiones de este proceso (sección IV).

II. DISEÑO DEL MOOC

El MOOC “*Introducción a la Programación con Java – Parte 1: Empezando a Programar en Java*” tiene una duración de cinco semanas de contenidos de aprendizaje y fue lanzado el 28 de abril de 2015, por lo que en el momento de escritura de este artículo todavía se encuentra en el periodo de impartición. El MOOC tiene previsto terminar el 30 de junio dando un periodo extra sobre las cinco semanas iniciales para que los alumnos, especialmente los que se incorporan a última hora, terminen las actividades de evaluación. El curso arrancó con más de 60.000 alumnos registrados de 190 países. Este MOOC es el primero de una serie de tres módulos (MOOCs) dedicados a enseñar los principios básicos de la programación utilizando como herramienta principal el lenguaje Java. El MOOC se ha desplegado en la plataforma edX y se imparte enteramente en inglés, aunque los vídeos en los que se presentan los contenidos teóricos están subtítulos en español, portugués y chino mandarín, además de en inglés. La carga de trabajo estimada del alumno es de 5-7 horas de trabajo a la semana.

A. Ciclo de trabajo

El equipo docente del MOOC lleva a cabo su diseño y despliegue en colaboración con la UTEID. Para ello se establece un ciclo de trabajo en el que el equipo docente se encarga de generar la estructura del curso en la plataforma *Studio Edge* [15] (donde *Studio* es la herramienta de autoría de edX y *Edge* es la instalación de pruebas de edX como paso a la migración definitiva a *edx.org*); graba los vídeos utilizando los equipos y salas de audiovisuales gestionadas por la UTEID; genera las actividades y laboratorios; y monta la versión final de los vídeos y ejercicios en *Studio Edge*. Por otro lado, la UTEID se encarga de producir las versiones finales de los vídeos (añadiendo transparencias, efectos y animaciones cuando procede); crea, corrige y gestiona los subtítulos asociados a cada vídeo; añade los componentes de discusión (foro de edX) relacionados con cada contenido y la pestaña correspondiente a redes sociales a utilizar en el curso (Twitter en este caso); migra de *Edge* a *edx* antes del comienzo del curso; y mantiene copias de seguridad del curso regularmente. La UTEID además da soporte adicional en ciertas tareas como la generación de algunos ejercicios especiales que requieren animaciones, o el despliegue de ejercicios con cierta complejidad en edX como los de evaluación entre pares.

Como sistema de compartición y gestión de contenidos se utiliza una herramienta propia de la UC3M llamada GEL [4]. GEL permite almacenar vídeos y publicarlos automáticamente en YouTube y al mismo tiempo en un servidor propio de la UC3M. Además se utiliza Google Drive para hacer un seguimiento del estado de cada uno de los materiales, compartir y editar guiones y transparencias de cada uno de los vídeos, compartir y editar ejercicios, y compartir y editar los mensajes a enviar a los alumnos.

B. Visión general y estructura

Los tres primeros pasos en el diseño y despliegue de un MOOC en edX son críticos ya que permiten promocionar el curso y captar potenciales alumnos. Estos son: la creación de una “*about page*” o *página de presentación* estándar con la información general del curso (descripción breve, objetivos, descripción del profesorado, carga semanal, fechas de impartición, idioma de impartición); la creación de un *video de presentación* atractivo sobre el curso de entre uno y dos minutos que se añade a la página de presentación; y la creación del “*syllabus*” o *ficha del curso* en la que se incluye al menos la descripción y objetivos detallados, la estructura del curso desglosada por semanas, el sistema de evaluación, el calendario del curso y se indica también cómo se llevará a cabo la comunicación del profesorado con los alumnos. En el caso de este MOOC se contrataron los servicios profesionales de una empresa externa a la universidad para el vídeo de presentación, y además se fue actualizando la versión del “*syllabus*” a medida que se iban diseñando las diferentes semanas del mismo.

El MOOC consta de cinco semanas con una estructura similar precedidas de una semana 0. La semana 0 se utiliza para incluir un par de vídeos de introducción y motivación relacionados con la temática del curso, una encuesta para conocer mejor a los alumnos y un componente de discusión para que estos puedan presentarse a sus compañeros. Cada semana (o sección en terminología edX) tiene una estructura similar y está dividida en subsecciones. Cada subsección tiene una o varias unidades y cada unidad puede tener uno o varios componentes (p.ej. un componente de vídeo, otro de texto y otro de discusión). Las subsecciones pueden clasificarse de la siguiente manera (ver Figura 1 para el detalle de las subsecciones para la primera semana):

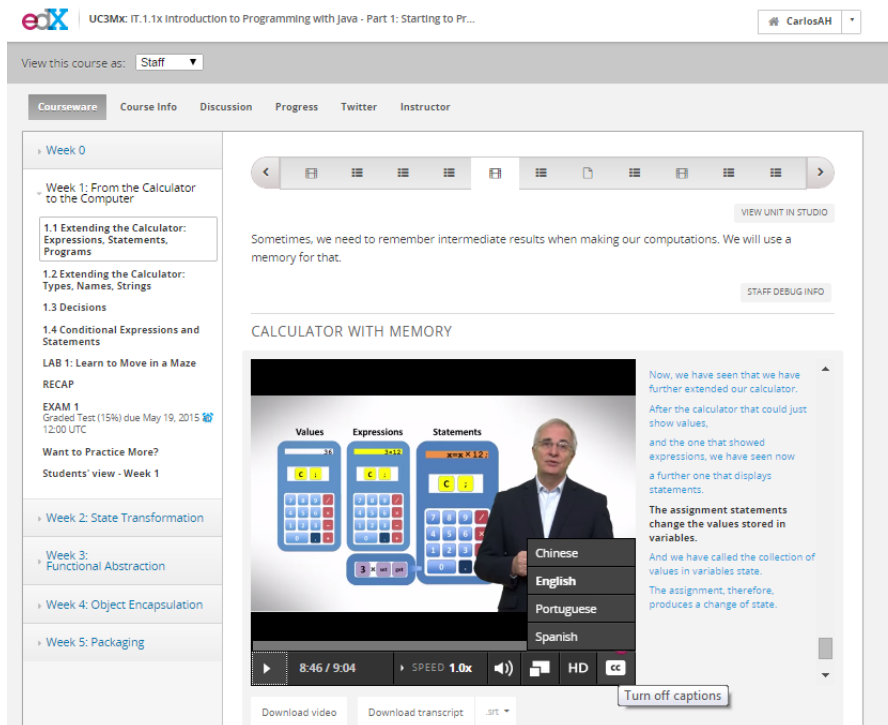


Fig. 1. Captura de pantalla del MOOC. En la barra izquierda se observa la estructura del curso (*Courseware*) y el detalle para la primera sección o semana (“*Week 1: From the Calculator to the Computer*”) con sus subsecciones. En la barra superior derecha se observan las diferentes unidades para el caso de la subsección 1.1 (“*Extending the Calculator: Expressions, Statements, Programs*”). En la parte inferior derecha se presenta uno de los vídeos (*Calculator with Memory*) con subtítulos en inglés, español, portugués y chino; tanto el vídeo como los subtítulos pueden descargarse.

- **Contenidos principales de la semana.** Estos se distribuyen en cuatro subsecciones (o cuatro cuartos) cada uno de ellos presentando y ejercitando conceptos relacionados. En la Figura 1 puede verse el caso de la primera semana con las subsecciones 1.1 a 1.4. Cada una de estas subsecciones contiene varias unidades en las que se incluyen vídeos (típicamente de 2 a 5 vídeos cortos por subsección) y múltiples ejercicios relacionados con los conceptos explicados en los vídeos. Estos ejercicios pueden desarrollarse sobre las herramientas propias de edX o sobre herramientas externas como se detalla más adelante. Además, puede haber explicaciones complementarias textuales sobre los conceptos explicados e incluso código que los alumnos tienen que descargar para trabajar sobre él.
- **Contenidos de laboratorio (Lab).** Todas las semanas incluyen varios ejercicios relacionados entre sí, agrupados bajo la subsección laboratorio, en la cual se ejercitan los contenidos aprendidos en la semana correspondiente. Las claves del diseño del laboratorio son las mismas que en las que se basan muchos juegos [16]: motivación voluntaria del alumno, fallar pronto y de una forma divertida; aplicar lo que se ha aprendido para alcanzar un objetivo. Estos tres principios se aplican en dos niveles de granularidad: en cada ejercicio individual; y en el laboratorio global, el cual sigue una estructura incremental a lo largo de las cinco semanas de tal forma que se van integrando diferentes ejercicios hasta conseguir un juego completo. La metodología aplicada en el laboratorio está basada parcialmente en la metodología PhyMEL (*Physical, Mental and Emotional Learning*) [17] según la cual al principio se hacen ejercicios cortos de entrenamiento que luego se combinan para conseguir un proyecto global y en el que se intentan evocar aspectos: físicos (p.ej. un muñeco que se mueve por un laberinto); emocionales (satisfacción por las recompensas a corto plazo y a largo plazo como conseguir superar los ejercicios individuales y el proyecto global); y mentales (comprender conceptos y algoritmos relacionados con la temática de cada semana). Por restricciones en los tipos de ejercicio propios de la plataforma, en el laboratorio se utilizan herramientas externas a edX, las cuales se detallan más adelante.
- **Recapitulación (Recap).** En esta subsección se proporciona un resumen textual con los conceptos más importantes que han aparecido a lo largo de la semana para que sirva de apoyo al trabajo del alumno. En esta subsección también se proporcionan las soluciones a los ejercicios más complejos planteados durante la semana, particularmente aquellos en los que se utilizan herramientas externas a edX.
- **Examen (Exam).** En esta subsección se plantea un examen con varios ejercicios de corrección automática sobre los temas tratados en la semana. Los alumnos tienen que realizar el examen en un periodo de tiempo delimitado (típicamente dos semanas después de la apertura de los contenidos de esa semana).
- **Ejercicios extra (Want to practice more?).** Se ofrece a los alumnos una subsección con un conjunto de ejercicios adicionales sobre esa semana para aquellos que quieran continuar trabajando y ejercitando todo lo explicado en la semana.
- **Visión de los estudiantes (Students' view).** Una vez finalizada la semana se ofrece a los alumnos en esta subsección un vídeo resumen de unos 5-8 minutos grabado por estudiantes del Colegio Americano de Madrid (16-17 años) que han trabajado los contenidos de la semana y que comentan informalmente sus impresiones en parejas o grupos de tres con un lenguaje más cercano al alumno medio que cursa este MOOC. En los vídeos, los estudiantes explican con sus propias palabras algunos de los conceptos más complejos de la semana. Además, comentan aquellas actividades o vídeos que más les han gustado y cuáles consideran que son mejorables. El objetivo principal de estos vídeos es conseguir que el alumno que participa en el MOOC vea reflejado su propio proceso de enseñanza-aprendizaje en los comentarios y reflexiones de estudiantes con un nivel semejante al suyo. Se pretende conseguir un aumento de motivación de los alumnos, además, de fomentar la reflexión del alumno acerca de los contenidos de la semana y de su propia evolución en el curso.
- **Evaluación entre pares (peer assessment).** En dos de las semanas (semanas 3 y 5) se incluye entre el examen y la subsección de ejercicios extra una subsección adicional con un ejercicio de evaluación entre pares, tal y como se explica más adelante, en el que los alumnos del MOOC deben escribir un pequeño fragmento de código en Java que es evaluado por sus propios compañeros.

C. Vídeos

Los vídeos son los elementos centrales de las explicaciones teóricas en este MOOC. En total se han creado 64 vídeos con contenido propio y desde cero, los cuales se distribuyen a lo largo de las cinco semanas del curso. Todos los vídeos tienen una duración mayor de 3 minutos y menor de 9 minutos, dependiendo de los contenidos a tratar. La mayor parte de los vídeos está en el rango de los 5-7 minutos, el cual es un criterio habitual de buenas prácticas en este tipo de cursos. Por ejemplo, la duración óptima de un vídeo para maximizar la motivación del alumno es de unos 6 minutos según edX [18], aunque ésta puede variar en función de la materia explicada.

Todos los vídeos están grabados en inglés, aunque se ofrecen con subtítulos en cuatro idiomas: inglés, español, portugués y chino mandarín. Para ello se ha realizado una colaboración con la empresa *3Play Media* a través de edX, la cual se ha encargado de proporcionar los subtítulos en inglés, portugués y chino mandarín, mientras que la UTEID se ha encargado de realizar las transcripciones al español. De esta forma se busca mejorar la experiencia del alumno de este MOOC, independientemente de su país de origen y lengua materna. Un paso importante y necesario en el ciclo de trabajo del subtítulo es realizar primero las transcripciones en inglés y que un profesor del curso se encargue de corregirlas manualmente, antes de que pueda pasarse a la fase de transcripción a los otros tres idiomas, ya que siempre se detectan algunos fallos de transcripción en la terminología propia del curso.

Todos los vídeos del curso se ofrecen en alta resolución y en baja resolución. Tanto los vídeos en alta y en baja resolución, como los subtítulos en los diferentes idiomas se ofrecen a los alumnos para su descarga. El hecho de ofrecer vídeos con baja resolución está motivado para facilitar que

aquellos alumnos que tengan un menor ancho de banda o cuyos proveedores de servicios de Internet impongan limitaciones en sus tarifas de descarga puedan también seguir cómodamente el curso. Los vídeos se alojan en YouTube y en un servidor propio de UC3M (para su descarga) y se enlazan desde edX. Además, edX se ha encargado de convertir los vídeos del MOOC para que puedan ser accesibles desde la aplicación móvil de edX, la cual está disponible para Android y iOS.

A lo largo de los diferentes vídeos aparecen varios de los profesores del curso, lo cual puede conllevar dificultades debido a los distintos estilos de cada uno de ellos. Sin embargo, se ha hecho un esfuerzo desde el diseño del curso para homogeneizar el estilo de los vídeos y mantener un mismo nivel de calidad. Por ejemplo, la mayoría de los vídeos están grabados en un estudio de grabación con un croma detrás y utilizan transparencias o animaciones de apoyo. Algunos vídeos utilizan únicamente capturas de pantalla (mostrando por ejemplo el funcionamiento de un trozo de código), por lo que no era necesario el uso del estudio de grabación, aunque siempre se ha utilizado un micrófono profesional para minimizar posibles ruidos y mejorar la calidad del audio. Uno de los problemas al grabar con croma y posteriormente eliminarlo en producción es que el color blanco del fondo puede dar una sensación de vacío al alumno que ve el vídeo. Por ello se ha incluido en todos los vídeos una “viñeta” elíptica que desvanece del color blanco al negro en los bordes del vídeo. Además, se ha incorporado en todos los vídeos un sonido característico hacia el final del mismo que da pie a un breve resumen de lo que se ha tratado en el vídeo.

El contenido de los vídeos propiamente dicho suele amenizarse con la utilización de anécdotas de la vida real (p.ej. la toma de decisiones en un programa ejemplificada a través de la decisión del profesor de salir de casa con un paraguas dependiendo de si llueve o no). Además, se utilizan objetos del mundo físico para acompañar los vídeos (en el ejemplo anterior un paraguas que el profesor lleva en el vídeo, o unas muñecas rusas que el profesor utiliza para explicar el concepto de recursión) y se cuida que el profesor tenga un tono agradable. Los vídeos van precedidos siempre de un breve texto en la plataforma, el cual sirve para resumir brevemente lo que se va a contar o plantear una pregunta de reflexión.

Además de los 64 vídeos de contenidos propios del curso se han generado: un vídeo de introducción; un vídeo de cierre y despedida de los profesores; y cinco vídeos con la visión de los estudiantes grabados con alumnos del Colegio Americano de Madrid. En la medida de lo posible se ha intentado ofrecer también estos vídeos en alta y baja calidad y con subtítulos en los distintos idiomas.

D. Actividades

El MOOC ha sido diseñado para fomentar la interacción del alumno con los materiales y enseñarle a programar progresivamente a través de múltiples actividades. De hecho en cada semana se han diseñado más de 50 actividades de diferente duración y dificultad las cuales forman parte de las subsecciones: contenidos principales, contenidos de laboratorio, examen y ejercicios adicionales. En el caso de las subsecciones de contenidos principales, las actividades se

entrelazan con los vídeos, de tal forma que ayudan a afianzar los conceptos que acaban de ser explicados. En el caso de los contenidos de laboratorio las actividades se enmarcan dentro de una historia denominada la “mansión encantada” y durante la cual el alumno tendrá que aprender a moverse por laberintos tomando decisiones y recurriendo a repeticiones de acciones a través de bucles, programar un algoritmo de “backtracking” (o vuelta atrás) para resolver un laberinto y aprender los principios de orientación a objetos creando personajes (objetos) que le ayuden en su misión en la mansión. En el caso de la subsección de ejercicios adicionales, éstos se plantean como retos para aquellos alumnos más ambiciosos o que ya tienen conocimientos previos y que por tanto necesitan menos tiempo para completar el resto de ejercicios.

Para diseñar las actividades del MOOC se han utilizado tanto herramientas propias de edX como herramientas externas a edX. Las herramientas externas estaban o bien integradas en edX (mediante código JavaScript, a través de un iFrame, o a través del estándar LTI [19] – *Learning Tools Interoperability* – el cual es soportado por edX), o bien requerían que el alumno se las descargase e instalase en su máquina. Para la evaluación sumativa de los alumnos únicamente se han utilizado herramientas propias de edX, debido a la dificultad de integrar las herramientas externas con el sistema de calificaciones de edX.

Dentro de las herramientas propias de edX destacan las de corrección automática. Entre ellas se han utilizado los siguientes tipos: problema de respuesta múltiple (*multiple choice problem*), problema de casillas de verificación (*checkbox problem*), problema de lista desplegable (*dropdown problem*), problema de entrada de texto (*text input problem*) y problema de arrastrar y soltar (*drag and drop problem*). De estos merece la pena destacar los dos últimos. Los problemas de entrada de texto deben usarse con mesura ya que son bastante conflictivos al requerir que el alumno proporcione una solución determinada y exacta tal y como se define en la configuración del ejercicio. Añadir espacios, utilizar mayúsculas y minúsculas indistintamente, o signos de puntuación puede conllevar que el sistema considere incorrecta una solución que puede ser perfectamente válida. Por otro lado, los problemas de arrastrar y soltar son problemas muy interactivos en los que el alumno tiene que arrastrar elementos desde una barra inferior a una imagen que aparece justo encima. El profesor define la parte de la imagen en la que ese elemento representa una solución válida durante la configuración de la actividad. Este tipo de actividades son costosas de diseñar, debido a la necesidad de definir manualmente el espacio de soluciones posibles, pero tienen una buena aceptación entre los alumnos debido a su interactividad. Un inconveniente importante de los problemas de arrastrar y soltar es que actualmente no funcionan bien en algunos dispositivos móviles como *tablets* y *smartphones*. La Figura 2 muestra un ejemplo (ya resuelto) de un problema de arrastrar y soltar en el que un alumno debe rellenar huecos en un código en Java utilizando los operadores que se le proporcionan en la barra inferior; algunos de los operadores deben utilizarse varias veces mientras que otros no deben utilizarse para resolver correctamente el problema.

CONDITIONALS (NON GRADED ACTIVITY) (1/1 point)

The body mass index (BMI) is a well-known number to classify people according to their body fat. It is calculated as the mass of the person (in kg) divided by the square of the height (in meters). The following code segment calculates the category of a person given his/her mass and weight. Complete the code using the operators that are shown below and taking into account the categories and BMI values shown in the table.

Note: `double` is another data type in Java like `int`, but that allows representing decimal values. We will see more about data types later in this course.

(This exercise may not be compatible with mobile devices: we recommend that you use a computer)

Category	BMI
Underweight	Less than 18.5
Normal weight	From 18.5 (inclusive) to 25 (exclusive)
Overweight	From 25 (inclusive) to 30 (exclusive)
Obese	From 30 on

```
double mass = 90;
double height = 1.90;
double bmi = mass / (height * height);
String category;

if (bmi < 18.5) {
    category = "underweight";
}
if (bmi >= 18.5 && bmi < 25) {
    category = "normal weight";
}
if (bmi >= 25 && bmi < 30) {
    category = "overweight";
}
if (bmi >= 30) {
    category = "obese";
}
```



Fig. 2. Ejemplo de problema de *drag and drop* (ya resuelto) donde el alumno debe rellenar los huecos en el código de la parte superior con los operadores proporcionados en la parte inferior.

Otra de las herramientas propias de edX utilizadas en este MOOC es la evaluación entre pares (*peer assessment – P2P*). En las dos actividades P2P incluidas en este curso se proporciona a los alumnos un enunciado de un código a resolver y se les da una “rúbrica” (o criterios) de evaluación detallada. Los alumnos suben su tarea resuelta a edX, el cual distribuye cada tarea a un número de compañeros determinado durante la configuración de la actividad (típicamente 3-5, pero solamente considerando aquellos que también han entregado la tarea). Para facilitar la evaluación entre pares es muy importante definir una “rúbrica” concreta, con las diferentes partes a evaluar claramente segmentadas, incluyendo los valores cuantitativos posibles para cada parte, y ejemplos para cada valor cuantitativo. Este tipo de actividades normalmente requieren dos plazos: un primer plazo para la entrega de la actividad y un segundo plazo para la revisión entre pares, por lo que es necesario dejar tiempo suficiente a los alumnos en ambos casos. En este MOOC las actividades de evaluación entre pares únicamente se han utilizado como parte de la evaluación sumativa del curso.

Además se han integrado varias herramientas externas desarrolladas en JavaScript por el equipo docente específicamente para el curso. Entre ellas destacan: varios tipos de calculadoras de complejidad variable (con operadores para números enteros, booleanos y ambos); un coche con control remoto; y un conversor de decimal a binario, incluyendo soporte a números negativos a través del sistema de representación en complemento a dos.

También se ha integrado la herramienta web de Google *Blockly* [20] (ver ejemplo en Figura 3), la cual permite crear juegos sencillos a través de una representación visual. Con *Blockly* el alumno ensambla bloques de pseudocódigo arrastrando y soltando desde un panel central hacia un panel en la parte derecha, y sin necesidad de conocer la sintaxis del

lenguaje de programación. *Blockly* se ha particularizado para generar una secuencia de laberintos de dificultad incremental en los laboratorios de la primera y segunda semana del MOOC adaptados a la temática de la mansión encantada.

Como herramienta de compilación (y ejecución) de código Java desde el navegador se ha integrado *Codeboard* [21] (ver ejemplo en Figura 4). Esta herramienta, integrada a través de LTI en edX, ha permitido la inclusión de numerosos ejercicios de apoyo, creados y salvados previamente por el profesorado, que el alumno puede cambiar, compilar y ejecutar de forma dinámica desde su navegador. La herramienta establece un tiempo máximo de ejecución para evitar problemas como bucles infinitos, muy comunes entre los alumnos que están aprendiendo a programar. Si el alumno comete un error o desea empezar de nuevo el ejercicio, basta con refrescar la página del navegador para encontrar otra vez el ejercicio en su estado inicial. En el caso de que el alumno quiera guardar los cambios realizados, puede abrir una cuenta en la plataforma; esta opción se les menciona a los alumnos, pero no es en absoluto obligatoria. La herramienta se ha usado como parte de los laboratorios a partir de la semana 2. Es importante comentar que esta herramienta externa no se usa para validar ni entregar código (aunque sería posible); sólo se usa para que los alumnos aprendan a través de la modificación (y creación) de pequeños programas.

A partir de la semana 4, y como parte de la programación orientada a objetos, se introduce la herramienta *Greenfoot* [22]. *Greenfoot* permite el desarrollo de juegos de forma simplificada y ha sido usado tanto para la elaboración de ejercicios relacionados con los contenidos principales de la semana, como de los laboratorios de las semanas 4 y 5. Esta herramienta debe ser descargada por el alumno e instalada en máquina, por lo que ha sido necesaria la inclusión de explicaciones acerca de la descarga e instalación *Greenfoot*. Al igual que con *Codeboard*, el profesorado ha creado pequeños ejercicios que los alumnos deben descargar y ejecutar. Además, se han realizado vídeos de apoyo sobre el uso de *Greenfoot* y como guía en la realización de las primeras actividades con esta herramienta (ver Figura 5).

MORE BRANCHING

We understand that branching may be a little bit tricky. Here you have another level so you can practice more. Take into account that in this level branches allow you to evaluate three different conditions. 7 blocks are allowed in this level but we are sure that you can do it with just 6.

(This activity may not be compatible with mobile devices: we recommend that you use a computer)



Fig. 3. Ejemplo de ejercicio con *Blockly* integrado en edX como parte del laboratorio de la semana 1 en el que los alumnos deben llegar al final de un laberinto utilizando bloques que representan movimientos, giros, bifurcaciones y bucles.

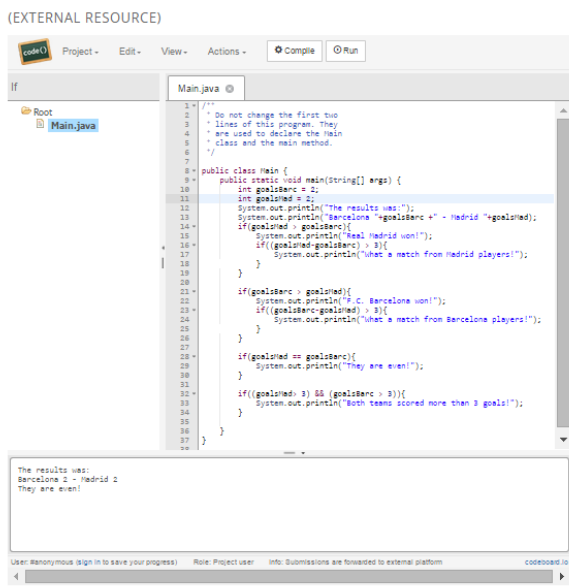


Fig. 4. Ejemplo de ejercicio con Codeboard integrado en edX como recurso externo a través de LTI.

E. Sistema de evaluación

El sistema de evaluación está basado exclusivamente en ejercicios realizados sobre herramientas propias de edX. Esta decisión se ha tomado para evitar depender en algo tan sensible de servicios ofrecidos por terceros ajenos a UC3M y edX. La evaluación sumativa consta de exámenes basados en ejercicios de corrección automática (5 exámenes), y de actividades más complejas de evaluación entre pares (2 actividades).

Cada semana los alumnos deben realizar un examen consistente en múltiples preguntas (entre 5 y 10), independientes entre sí, sobre los temas tratados durante la semana. Cada examen tiene un peso del 15% de la nota final, con lo que el peso total de los exámenes es de un 75% de la nota final. En la elaboración de estos ejercicios se han tenido en cuenta las limitaciones actuales de la plataforma. Por ejemplo, usando problemas de arrastrar y soltar (*drag and drop*) con múltiples opciones, si el alumno se equivoca en un solo campo, la pregunta es evaluada por la plataforma con un cero. Debido a esta limitación, se decidió minimizar el uso de preguntas de este tipo en los exámenes.



Fig. 5. Ejemplo de ejercicio realizado con Greenfoot (izquierda) en el que los alumnos desarrollan un pequeño juego de carreras de coches. Explicación del funcionamiento del ejercicio y del código a través de uno de los videos.

Además, se han programado dos actividades de evaluación entre pares, una en la semana 3 (10% de la nota final), justo antes de empezar con programación orientada a objetos, y otra en la semana 5 (15% de la nota final), al finalizar el curso. La actividad de evaluación entre pares de la semana 3 plantea un problema específico de los contenidos de esa semana, mientras que la actividad de la semana 5 plantea un problema global de todo el curso. En este tipo de actividades tanto el enunciado como la rúbrica de corrección se han cuidado especialmente. Además, al estar orientado el MOOC a neófitos en la programación, es muy probable que sea la primera vez que se enfrenten a la elaboración de un código desde cero, por lo que se han planteado actividades muy guiadas.

F. Comunicación con los alumnos

La comunicación con los alumnos se ha estructurado de la siguiente manera:

- **Mensaje semanal.** Al inicio de cada semana, y coincidiendo con la liberación nuevos materiales, se envía un mensaje (que reciben los alumnos en el correo que han proporcionado a la plataforma) que presenta un resumen de las actividades y conceptos más importantes que se cubrirán durante esa semana. Además, el mensaje sirve para recordar aspectos logísticos: plazos de entrega; la posibilidad de bajarse no sólo los vídeos y subtítulos, sino también las transparencias en PDF; la necesidad de consultar regularmente el “syllabus”, etc. A partir de la segunda semana los vídeos del Colegio Americano se anuncian también en el mensaje semanal. Se pretende con este mensaje motivar a los alumnos e invitarles a entrar en la plataforma para ver estos contenidos.
- **Información acerca del curso (Course Info).** Los mensajes semanales son publicados también en la página de información acerca del curso dentro de la plataforma. Esta página, además, contiene enlaces al “syllabus”, una guía de cómo usar los foros y un convertidor de hora local a UTC (que es el sistema seguido en el curso).
- **Uso de foros dentro de edX.** Se ha decidido usar como canal de comunicación principal con los alumnos los foros proporcionados por la propia plataforma.
- **Uso de twitter.** Para el curso se ha creado una cuenta específica en twitter (@javaedxuc3m) y se pide a los alumnos que usen el hashtag (#javaedxuc3m) para mantener conversaciones en esta red social.

III. ALGUNAS LECCIONES APRENDIDAS

Esta sección presenta algunas de las lecciones aprendidas por el equipo docente durante el proceso de diseño y puesta en marcha de este MOOC. Estas lecciones cubren tanto el proceso de diseño y puesta en marcha de un MOOC en general como la particularización para el caso de edX.

Una de las principales lecciones aprendidas se refiere a la *gestión del tiempo*. Es muy fácil subestimar la dedicación necesaria del profesorado en la creación de un MOOC. Es por ello necesario hacer una planificación concienzuda del tiempo a dedicar desde antes de comenzar la etapa de diseño (y ajustarse a ella), dejando tiempo suficiente para discutir el diseño del curso y generar los materiales, pero también para el testeado del curso y la subtitulación de los vídeos. Respecto a los vídeos es además muy importante tener en cuenta el precio tan costoso que supone realizar cambios a posteriori (volver a producir, cargar en la plataforma, ajustar el subtítulo...) por lo que deben revisarse concienzudamente antes de darlos por definitivos.

Este MOOC ha sido producido por un equipo de ocho docentes por lo que algunas de las lecciones aprendidas más valiosas se refieren a la *coordinación* del trabajo. A la hora de buscar la eficiencia es importante establecer reuniones específicas cada semana (por tareas o módulos) con las personas implicadas y que alguien asuma el rol de coordinador. Tener reuniones generales es importante al principio y al final para tener una visión global, pero en el día a día, y especialmente si la dedicación del profesorado no puede ser completa, es recomendable tratar de compartimentar. Además, es importante partir de un esquema inicial lo más detallado posible del contenido que se quiere ofrecer. Esto evita repetir ejercicios al repartir el trabajo y ayuda a guardar la coherencia a la hora de tratar ciertos conceptos. Es necesario hacer un buen reparto de ejercicios en grupos pequeños de trabajo y con revisiones cruzadas de personas que han desarrollado contenidos relacionados para minimizar los errores en dichos ejercicios. En esta línea, también es necesario tener un plan de contingencia ante posibles bajas del profesorado. En este curso se produjo la baja de tres de los nueve profesores iniciales de la universidad donde se impartía el MOOC y, aunque dos de ellos siguieron activamente participando en el MOOC, su grado de implicación no pudo ser el mismo que habían estimado inicialmente, generando una carga adicional para el resto de integrantes del equipo.

En cuanto a las *actividades* planteadas durante el curso cabe destacar la necesidad de tener en cuenta la importancia de disponer de herramientas fáciles de usar que permitan dotar al curso de contenidos relacionados con la materia que se quiere impartir. Por ejemplo, edX no cuenta a día de hoy con herramientas nativas que permitan trabajar con código Java (editar, compilar y ejecutar). Por esto motivo ha sido necesario integrar herramientas externas como *Codeboard*, o *Greenfoot*. Utilizar herramientas que pueden funcionar desde el navegador sin necesidad de instalar entornos pesados puede contribuir a disminuir el porcentaje de abandonos en cursos de iniciación como éste. Sin embargo, hay que tener en cuenta también las limitaciones de algunas herramientas a la hora de funcionar en ciertos dispositivos y navegadores incluyendo Blockly o Codeboard pero también los ejercicios de arrastrar y soltar (nativos en edX).

El hecho de plantear un *laboratorio*, proyecto o caso de estudio longitudinal al MOOC, es en general positivo, ya que permite reforzar conceptos y que los alumnos vayan encajando las piezas, construyendo soluciones incrementales, lo que suele tener un efecto positivo en la motivación de los alumnos. Por el contrario, una aproximación de estas características requiere una gran coordinación entre los diferentes ejercicios relacionados y cualquier modificación puede afectar transversalmente a todas las semanas.

Otros de los aspectos más importantes en un MOOC es la *componente social*. Es necesario dedicar tiempo durante la etapa de diseño a preparar una estrategia social para el curso. En esta estrategia se determinan las herramientas a utilizar (foros, Twitter...) pero también los roles de cada componente del equipo. El hecho de no haber planteado una estrategia social a tiempo puede suponer que se desaprovechen algunos de esos canales de comunicación con los alumnos. También es importante conocer las limitaciones que tienen cada uno de esos canales de comunicación. Por ejemplo, en el caso de edX, uno de los principales problemas encontrados en la comunicación con los alumnos en el foro de la propia plataforma es la desorganización de los diferentes hilos de discusión. En este sentido es muy difícil seguir una conversación sobre un tema, pues los hilos de discusión no pueden agruparse de forma sencilla. Para amortiguar este problema se decidió asociar un componente de discusión a cada unidad en edX, identificándolo mediante el título de la semana y el título de la unidad. Así, se ha conseguido, al menos, detectar problemas relacionados con vídeos y ejercicios concretos. Por otra parte, en edX el manejo y gestión del foro en sí mismo es bastante complejo (tanto para alumnos como para profesores). Por ejemplo, el profesor no puede reordenar los hilos de discusión para una mayor comprensión por parte de los alumnos. Además, cada vez que se abre un hilo, se mueve la barra de navegación lateral lo cual dificulta la lectura del foro. Esto contribuye a que haya alumnos que no usen de forma adecuada el foro, abriendo hilos a preguntas ya resueltas, y que éste (exceptuando la organización mínima dada mediante el uso de títulos decidida por el profesorado) esté bastante desorganizado. Todo esto provoca que el mantenimiento de este canal de comunicación propio de la plataforma sea muy costoso.

Otra lección aprendida consiste en valorar convenientemente el esfuerzo de los *contenidos y actividades adicionales*. En el caso de este curso se proporcionaban por ejemplo vídeos grabados con alumnos del Colegio Americano de Madrid como complemento. Sin embargo, la logística detrás de la grabación de esos contenidos fue costosa e involucró desplazamientos de un equipo de cámara y sonido más el profesor durante varias sesiones, el visionado de horas de grabación para resumir los aspectos más relevantes y finalmente la producción final de los vídeos.

Por último, es importante contar con una *unidad de apoyo* como la UTEID a la hora de realizar un MOOC. De esta forma el profesorado puede abstraerse de muchos de los detalles técnicos de la creación de contenidos audiovisuales, para centrarse en añadir valor y contexto a los contenidos creados. Además, una unidad como la UTEID puede realizar un seguimiento periódico del estado del curso para intentar mantener la planificación inicial, y asumir la comunicación con el proveedor de la plataforma (edX en este caso).

IV. CONCLUSIONES

Este artículo ha presentado el diseño del MOOC “Introducción a la Programación con Java – Parte 1: Empezando a Programar en Java”, incluyendo el ciclo de trabajo utilizado en la UC3M, la visión general del curso, el diseño de los vídeos y actividades, el sistema de evaluación y la componente social del curso. Este MOOC forma parte de una serie de tres módulos dedicados a introducir la programación utilizando el lenguaje Java. Para ello en este primer módulo se ha utilizado una aproximación *bottom-up*, desde la implementación más básica basada en instrucciones y programación imperativa hasta el diseño de alto nivel basado en el paradigma de orientación a objetos. Queda también pendiente reutilizar parcialmente los materiales generados en este MOOC para realizar una segunda aproximación de este primer módulo siguiendo una aproximación *top-down*, desde el diseño de alto nivel hasta la implementación de bajo nivel, para permitir a los alumnos elegir la aproximación a través de la cual desean aprender los principios de la programación utilizando el lenguaje Java, ya que no existe un consenso generalizado en la comunidad con respecto a estas dos aproximaciones.

Además, en este artículo se han recogido algunas de las lecciones aprendidas por el profesorado durante el diseño y puesta en marcha de este MOOC. Estas lecciones aprendidas se refieren: a la necesidad de hacer un uso adecuado del tiempo de dedicación al MOOC; a la necesidad de preparar una buena estrategia de coordinación para la elaboración y testeo de los contenidos generados, así como de un plan de contingencia por si se producen bajas en el equipo docente; a la necesidad de fomentar la interactividad en el curso a través del planteamiento de numerosas actividades, las cuales pueden utilizar herramientas nativas de las plataformas o herramientas de terceros (con los problemas derivados que esto puede conllevar); a la necesidad de plantear proyectos o laboratorios con una cierta continuidad a lo largo del curso; a la necesidad de gestionar de una forma organizada los canales de comunicación con los alumnos; a la necesidad de valorar el esfuerzo que supone generar recursos adicionales de apoyo al curso; y a la necesidad de buscar apoyo especializado dentro de la universidad donde se va a llevar a cabo el curso.

AGRADECIMIENTOS

Este artículo ha sido parcialmente financiado por la Comunidad de Madrid a través de la Red de Excelencia eMadrid (S2013/ICE-2715). Este trabajo también ha recibido financiación por parte del Ministerio de Economía y Competitividad a través de los proyectos EEE (TIN2011-28308-C03-01) y RESET (TIN2014-53199-C3-1-R) y las Ayudas para contratos para la Formación Posdoctoral 2013 (FPDI-2013-17411). Nos gustaría agradecer también especialmente a la Unidad de Tecnología Educativa e Innovación Docente (UTEID) de la UC3M por contribuir en el diseño y despliegue del MOOC y a edX, así como a Diego Villarán, Jonathan Almodóvar y Raquel M. Crespo por su colaboración en la realización de diferentes ejercicios del MOOC, y al Colegio Americano de Madrid y particularmente a Joaquín, Elisa, Nico, Mylo, Carlos y Eduardo por sus aportaciones en los vídeos de la visión de los estudiantes.

REFERENCIAS

- [1] D. J. Skiba, “Disruption in Higher Education: Massively Open Online Courses (MOOCs)”, *Nursing Education Perspectives*, vol. 33, n. 6, pp. 416-417, 2012.
- [2] A. McAuley, B. Stewart, G. Siemens, D. Cormier, “The MOOC Model for Digital Practice”, 2010. Último acceso mayo 2015: http://www.elearnspace.org/Articles/MOOC_Final.pdf
- [3] A. Fox, “From MOOCs to SPOCs”, *Communications of the ACM*, vol. 56, n. 12, pp. 38-40, 2013.
- [4] C. Delgado-Kloos, P. J. Muñoz-Merino, M. Muñoz-Organero, C. Alario-Hoyos, M. Pérez-Sanagustín, H. A. Parada G., J. A. Ruipérez, J. L. Sanz, “Experiences of Running MOOCs and SPOCs at UC3M”, *Proceedings of the 2014 IEEE Global Engineering Education Conference, EDUCON 2014*, 884-891, Estambul, Turquía, 2014.
- [5] P. J. Muñoz-Merino, E. Méndez Rodríguez, C. Delgado Kloos, “SPOCs for Remedial Education: Experiences at the Universidad Carlos III de Madrid”, *Proceedings of the 2nd MOOC European Stakeholders Summit, EMOOCs 2014*, 271-276, Lausana, Suiza, 2014.
- [6] edX. Último acceso mayo 2015: <https://www.edx.org>
- [7] MiriadaX. Último acceso mayo 2015: <https://www.miriadax.net>
- [8] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, H. A. Parada G., M. Muñoz-Organero, “Delving into participants’ profiles and use of social tools in MOOCs”, *IEEE Transactions on Learning Technologies*, vol. 7, n. 3, pp. 260-266, 2014.
- [9] UTEID, Unidad de Tecnología Educativa e Innovación Docente. Último acceso mayo 2015: <http://www.uc3m.es/uteid>
- [10] UTEID, “Guía de buenas prácticas para profesores para la creación de vídeos educativos”, UC3M, 2014. Último acceso mayo 2015: http://portal.uc3m.es/portal/page/portal/biblioteca/UTEID/UTEID_GuiaCreacionVideosEdu-v.5.pdf
- [11] UTEID, “Guía metodológica para la planificación, diseño e impartición de MOOCs y SPOCs”, UC3M, 2014. Último acceso mayo 2015: http://portal.uc3m.es/portal/page/portal/biblioteca/UTEID/UTEID_GuiaMetodologicaMOOC%20v3Octubre2014.pdf
- [12] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, I. Gutiérrez-Rojas, D. Leony, H. A. Parada G., “Designing your first MOOC from scratch: recommendations after teaching ‘Digital Education of the Future’”, *Proceedings of the 2nd MOOC European Stakeholders Summit, EMOOCs 2014*, 172-177, Lausana, Suiza, 2014.
- [13] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, P. J. Muñoz-Merino, “Recommendations for the design and deployment of MOOCs: Insights about the MOOC Digital Education of the Future deployed in MiriadaX”, *Proceedings of the 2nd International Conference on Technological Ecosystems for Enhancing Multiculturality, TEEM 2014*, 403-408, Salamanca, España, 2014.
- [14] Introduction to Programming with Java - Part 1: Starting to Program in Java. Último acceso mayo 2015: <https://www.edx.org/course/introduction-programming-java-part-1-uc3mx-it-1-1x>
- [15] Studio Edge (edX). Último acceso mayo 2015: <https://studio.edge.edx.org>
- [16] J. McGonigal, “Reality is Broken: Why Games Make us Better and How they Can Change the World”, United States: Penguin Group, 2011.
- [17] C. Fernández-Panadero, C. Delgado Kloos, “PhyMEL. A Framework to Integrate Physical, Mental and Emotional Learning in Meaningful Experiences and Multidimensional Reports”, *3rd European Immersive Education Summit*, 203-209, Londres, Reino Unido, 2013.
- [18] P. Guo, “Optimal Video Length for Student Engagement”, edX, 2013. Último acceso mayo 2015: <https://www.edx.org/blog/optimal-video-length-student-engagement>
- [19] IMS Global Learning Consortium. Learning Tools Interoperability. Último acceso mayo 2015: <http://www.imsglobal.org/lti/index.html>
- [20] Blockly. Último acceso mayo 2015: <https://developers.google.com/blockly/>
- [21] Codeboard. Último acceso mayo 2015: <https://codeboard.io/>
- [22] Greenfoot. Último acceso mayo 2015: <http://www.greenfoot.org/>

Simtools: Teaching Networks with Virtual Scenarios

Juanjo Alins, Jose L. Muñoz, Jorge Mata, Oscar Esparza
 Departamento de Ingeniería Telemática,
 Universitat Politècnica de Catalunya (UPC)
 Jordi Girona 1-3, Campus Nord UPC. 08034 Barcelona.

Resumen—In this paper we present *simtools*, a software to create virtual networks of UML machines which is based on UML and VNUML. *Simtools* allows easily building virtual scenarios, and it is a quite simple and straightforward tool for teaching networking. In fact, we are using it in some of our subjects related to networking at the Network Engineering Department at the Universitat Politècnica de Catalunya (UPC). Additionally, we compare *simtools* with other systems and briefly describe our teaching experience with *simtools*.

I. INTRODUCTION

Let us consider the scenario of Figure 1. If you want to implement the previous topology, you need a certain investment for creating the network with real equipment: switches, cables, routers, hosts, servers etc. You also need to implement the physical configuration of the devices: connect cables, etc.

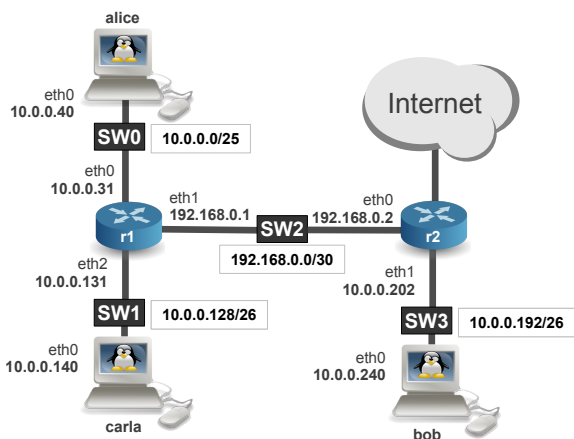


Fig. 1. An Example Scenario

Virtualization is a methodology for dividing the resources of a physical computer (**phyhost**) into multiple virtualized operating systems (OS). In a virtual network everything is implemented (real) except that hub/switches are software processes. With our tool for building virtual networks, we will just need:

- A **single computer** (called physical host or hypervisor) with the virtualization software installed.
- **Text files defining network topologies (scenarios).**

In the following sections we provide a brief introduction to virtualization and describe *simtools*, which is a tool based on VNUML [1], [2] for easily building virtual scenarios. *Simtools* is used in some of our subjects related to networking at the Network Engineering Department at the Universitat Politècnica de Catalunya (UPC). Finally, we compare *simtools* with other systems and briefly describe our teaching experience with it.

II. TYPES OF VIRTUALIZATION

There are several kinds of virtualization techniques for creating a virtual OS. These techniques provide similar features but differ in the degree of abstraction and the methods used [3], [4].

- **Hardware Emulation or Virtual machines (VMs).** This approach allows an hypervisor (**phyhost**) to run an arbitrary guest operating system. The guest OS is not modified and it is not aware that it is not running over real hardware. The main issue with this approach is that some OS instructions require to be in supervisor mode and this causes problems since the guest OS is being executed in the user space of the hypervisor. As a result, we need a virtual machine monitor (VMM) in the hypervisor to analyze executed code and to make it safe on-the-fly. This VMM is part of the “virtualization middleware”.
- **Paravirtualization.** In this virtualization approach most of the work of the VMM is implemented in the guest OS code, which is modified to avoid the use of privileged instructions. The paravirtualization technique also enables running different OSs on a single server, but requires them to be ported, i.e. guest kernels must “know” that they are running in a user space of an hypervisor.
- **Virtualization on the OS level, a.k.a. containers virtualization.** This technique shares a kernel for several virtual OS called “containers”, where each container has an isolated and secure environment. With containers you can even run different distributions but the kernel is shared among containers.

The previous three techniques differ in complexity of implementation, OS support, level of access to common resources and performance in comparison with an standalone server. In particular, hardware emulation has a wider scope of usage (many OS), but the poorest performance. Paravirtualization has a better performance than hardware emulation, but can support fewer OSs because these OS have to be modified. Containers virtualization has by far the best performance and scalability compared to the other two, but imposes that all the containers share the same kernel. Figure 2 shows a picture of the different virtualization types.

In Figure 2, we must point out that regarding containers, we show an example in which the hypervisor kernel is who implements the container technology (this is true for technologies like LXC and Docker). However, there are other historical container technologies (e.g. OpenVZ) which used special kernels over the hypervisor kernel to implement containers.

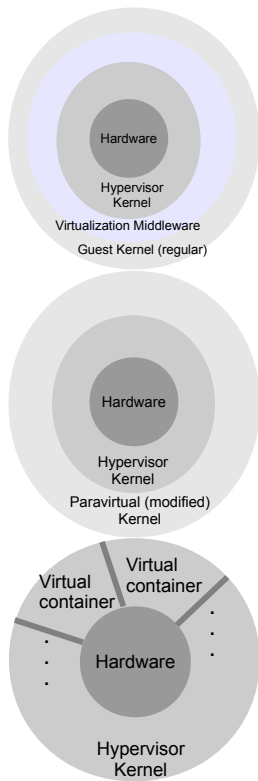


Fig. 2. Types of Virtualization

III. A VIRTUALIZED ENVIRONMENT

Virtualized environments have three basic elements (see Figure 3):

- **Hypervisor or Physical Host (phyhost).** This is the hardware, the operating system and any other software needed to run the virtual OS (guests).
- **Guests.** These are the virtual OS running over the **phyhost**. A guest might be a traditional OS running just like if it was on a real host. To do so, the host emulates all the system calls for hardware. This makes the guests feel like if they were in a real computer.
- **Virtual switches.** The virtual network is composed of virtual switches that connect guests like in physical networks. As an additional feature, the **phyhost** can provide connectivity for its guests, allowing them to exchange traffic with other physical networks and even with the Internet.

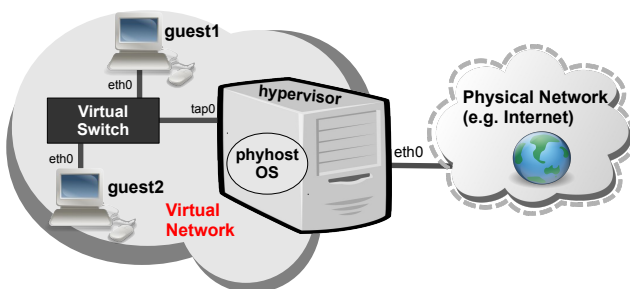


Fig. 3. Virtualization: a physical host and several guests

The guests are usually accessible with a CLI (Command Line Interface), a GUI (Graphical User Interface) or by some network service like SSH. On the other hand, the **phyhost** can have a virtual network interface that can be connected to the virtual switch. In Linux, we have several types of virtual interfaces like TUN/TAP and Veth.

IV. SIMTOOLS

Simtools is a tool for creating and running virtual environments of guests that use User Mode Linux (UML) kernels [5]. UML is a type of paravirtualization to run Linux guests over an unmodified Linux kernel that serves as hypervisor. Simtools provides several pre-configured scenarios using a modified version of a tool called Virtual Network User Mode Linux (VNUML) [1], [2]. VNUML allows us to easily define and run virtual networks using UML Kernels. In particular, VNUML provides us a language and a parser to create these scenarios. In this section, we explain how to install simtools and how its main scripts and configuration files work.

Currently, simtools is only packaged to be installed in Debian-based Linux systems. In particular, installations have been tested using Ubuntu 12.04/14.04 and Debian 7. To install simtools in Ubuntu you have to execute the following commands.

First, to add our repository to your list of APT repositories:

```
$ wget -O - http://sertel.upc.es/~vanetes/simtools/keyFile \
| sudo apt-key add -
$ echo "deb http://sertel.upc.es/~vanetes/simtools /" \
| sudo tee /etc/apt/sources.list.d/simtools.list
```

Then, type the following command to update the software lists:

```
$ sudo apt-get update
```

If your Linux is a **64-bit system**, install the package `libc6:i386`

```
$ sudo apt-get install libc6:i386 -y --force-yes
```

Finally, to install simtools:

```
$ sudo apt-get install simtools-meta-PACK -y --force-yes
```

Where PACK is equal to some of our subjects (`fxt`, `sat` or `tcgi`) or `*` to install all of them. You can repeat the previous steps if the software is not installed correctly at the first time, for example, if the network connection is lost.

A. Initial Configuration

Copy the configuration file of screen into your home directory:

```
$ cp /usr/local/share/doc/simtools/screenrc.user ~/.screenrc
```

You can also create a `vnuml` profile on your `gnome-terminal` (Edit- ζ Profiles- ζ New) with different colors to clearly differentiate the terminals of guests.

There are several important directories for simtools:

- **/usr/share/doc/upc/PACK.** Simtools installs documentation related with a subject and its scenarios in this directory. Remember that `PACK=fxt`, `sat` or `tcgi`.
- **/usr/share/vnuml/kernels.** Simtools stores the UML kernels of the guests in this directory.
- **/usr/share/vnuml/filesystems.** Simtools stores the filesystems of the guests in this directory.

- **/usr/share/vnuml/scenarios.** Simtools stores the definitions of the scenarios of in this directory. By default simtools searches predefined scenarios in this directory. If you want to use another directory, you can set the variable DIRPRACT.
- **\$HOME/vnuml.** Simtools stores data about executed scenarios in this directory (local changes).

You can update the software (scenarios, filesystems, documentation etc.) executing the following commands:

```
$ sudo apt-get update
$ sudo apt-get install simtools-meta-PAK -y --force-yes
```

Where PAK=fxt, sat, tcgi or *

B. Other Ways of Obtaining Simtools

In <http://bohr.upc.es/Networks> we provide updated information about simtools and also other ways of obtaining simtools.

We also provide a **Virtual Machine** with everything already installed on it that can run with hardware virtualization (e.g. with VirtualBox [6]). In this case, it is very important that you check that your processor supports hardware virtualization and that you activate this feature in your BIOS. You can check this with the following command:

```
$ grep -E "(vmx|svm)" --color=always /proc/cpuinfo
```

If nothing is displayed after running that command, then your processor does not support hardware virtualization, and you will not be able to use our virtual machine fluently.

Another possibility is to boot from an USB Stick with Ubuntu and simtools installed on it. Currently we have an image for a 16GB pendrive. Then, when booting, select in your system too boot from the USB stick (most modern computers support this disabling UEFI boot). To create the USB stick follow the instructions in our web site.

If you want to uninstall simtools type:

```
$ sudo apt-get remove simtools*
```

C. Start and Stop Scenarios

simctl is the main command to manage simtools. If you execute simctl without parameters, you will obtain the list of possible scenarios that you can run (by default looking at /usr/share/vnuml/scenarios). It is important to execute simctl always with your regular user (never as root). This is just an example of the output of simctl without parameters:

```
phyhost$ simctl
simctl ( dns-basic | fwnat | icmp-class | icmp | ip-routing-abc |
ip-routing | ip-subnetting | iptunnel | multicast |
netapps-basic | ospf-basic | rip | switching-vlan | tcp | www ) OPTIONS

OPTIONS
start           Starts scenario
stop            Stops scenario
status          State of the selected simulation
                (running | stopped)
vms             Shows vm's and tty's from simulation
labels [vm]     Shows sequence labels for ALL vm's or for vm
exec label [vm [vm [ ... ]]] Exec label in the vm's where label is defined
                or exec the label only in the vm list
netinfo         Provides info about network connections
get [-t term_type] vm [pts] Gets a terminal for vm
                term_type selects the terminal
                (xterm | gnome | kde | local)
                pts is an integer to select the console
forcestop       Destroys a simulation scenario
```

The output of simctl in this example tells us that it has located several scenarios (dns-basic, fwnat, etc). The previous output also shows us all the possibilities that simctl

provides us to manage the scenario. These possibilities are explored in the following sections. In addition, you can use the “TAB” key to auto-complete the commands and options available at each moment.

To start a particular scenario, you must execute simctl in the **phyhost** with the name of the selected scenario and use the start option. For example:

```
phyhost$ simctl ip-routing-abc start
.....
Total time elapsed: 123 seconds
phyhost$
```

Please, be patient because it might take some time to complete the starting process (this might take up to several minutes). Finally, the command ends indicating the time taken to start the scenario and we get the console prompt again. At this moment, all the virtual machines and their respective interconnections with virtual switches have been created.

After this, it is possible to check if the virtual interfaces have been created in the **phyhost** by typing ifconfig -a. We call these virtual interfaces SimNetX.

On the other hand, when you wish to stop the simulation, you can type the following:

```
phyhost$ simctl ip-routing-abc stop
.....
Total time elapsed: 17 seconds
phyhost$
```

Usually stopping a simulation is faster than starting it.

Important:

- You cannot start two simulations at the same time on the same phyhost.
- Remember to always stop your simulation before shutting down your phyhost.

A machine might spent some time while booting. Then, it might appear a message telling us to retry, continue or abort. **Type always retry (r).**

However, if a simulation never starts or stops, to clear the system, type CRL+c and then:

```
phyhost$ simctl simulation_name stop
phyhost$ simctl simulation_name forcestop
```

If your scenarios are not starting, try with:

```
phyhost$ simctl forcestop
```

And reboot **phyhost**.

The forcestop option kills all the “linux“ processes (UML kernels) and removes the directory .vnuml at the user’s home. After rebooting the system you should be able to start the simulation again.

On the other hand, you should never run two different simulations at the same time. If by mistake you start two simulations do:

```
phyhost$ simctl simulation_name1 stop
phyhost$ simctl simulation_name2 stop
phyhost$ simctl forcestop
```

Finally, you should never use the superuser “root“ to execute simctl. If by mistake you start a simulation with the root user, you must clear the system and start it again using your user:

```
phyhost$ sudo -s
phyhost# simctl simulation_name stop
phyhost# simctl forcestop
phyhost# exit
```


D. Using Virtual Machines

Once we have a simulation running, we can list the available virtual machines and control them by using `simctl`. Following the same example, let us assume that we have the `ip-routing-abc` scenario already running. At this moment, we can use the `vm` option to list the virtual machines (UML guests):

```
phyhost$ simctl ip-routing-abc vms
Virtual machines from ip-routing-abc:
num  vms  enabled tty's Id
1    alice  0
2    r1     0 1
3    r2     0
4    bob    0
5    carla  0
```

As we can observe, the machine `r1` has two consoles enabled.

The `get` option of `simctl` allows us to list and access to the virtual machines consoles. If you run the `get` option without parameters, you obtain information about the state of virtual machines (“Running” or “Not Running”) and the possibility to access their command consoles.

```
simctl ip-routing-abc get
alice      Running  -----
r1         Running  -----
r2         Running  -----
bob        Running  -----
carla     Running  -----
```

In the example shown, the dashed lines (- - - - -) indicate that all virtual machines have enabled consoles but that we have not yet accessed to any of them.

To access to the console of a virtual machine:

```
phyhost$ simctl ip-routing-abc get alice
```

The `get` option can have an argument (`-t`) to indicate what type of terminal you want to use (it requires that the selected terminal emulator is already installed in the system). You can also define a particular terminal as the default for your system in your preferences file “`simrc`”.

The terminal can be `xterm` (classic in X11), `gnome` (terminal from GNOME) or `kde` (Konsole from KDE).

For example, to get a `gnome-terminal` for `bob` virtual machine, you can type:

```
phyhost$ simctl ip-routing-abc get -t gnome bob
```

Once you have accessed to the console of the virtual machines `alice` and `bob`, you can type:

```
phyhost$ simctl ip-routing-abc get
alice      10681.0  (Attached)
r1         Running  -----
r2         Running  -----
bob        10777.0  (Attached)
carla     Running  -----
```

“Attached” indicates that there is already a terminal associated with the command console of `alice` virtual machine. If you close the `alice` terminal then the terminal state is “Detached”.

Finally, you can also access to the two consoles of `r1` using the following commands:

```
phyhost$ simctl ip-routing-abc get r1 0
phyhost$ simctl ip-routing-abc get r1 1
```

Using the syntax “`simctl simname netinfo`” you can access to the connection topology of virtual machines. For example:

```
phyhost$ simctl ip-routing-abc netinfo
UML      IFACE      NET
HOST     SimNet0    Net0
HOST     SimNet1    Net1
HOST     SimNet2    Net2
HOST     SimNet3    Net3
-----
alice    eth0       Net0
bob      eth0       Net3
carla    eth0       Net1
r1       eth0       Net0
r1       eth1       Net2
r1       eth2       Net1
r2       eth0       Net2
r2       eth1       Net3
```

The output is the topology defined in the XML file. This command can be useful to detect mistakes in the topology configuration. It is also worth to mention, that the “`netinfo`” option uses information directly obtained from the virtual machines (not from the VNUML file) when the simulation is running.

E. Managing and Executing Labels

VNUML allows you to define a set of actions to be executed on virtual machines while running a simulation scenario. VNUML uses labels specified in the “`seq`” attributes of the `<exec>` and `<filetree>` tags to associate a label name to a set of actions. Labels allow you to easily distinguish a certain set of actions from another set of actions. The assignment of label names to the associated actions takes place in the VNUML specification file, in particular, in the XML element that defines each virtual machine. The `simctl` wrapper has two options, “`labels`” and “`exec`”, to facilitate the management and execution of labels. To show the labels:

```
phyhost$ simctl example-scenario labels
uml1 : reset_ips
uml2 : reset_ips enable_forwarding disable_forwarding
uml3 : reset_ips
```

As shown in the output of the previous command, with the “`labels`” option we obtain the list of defined labels per virtual machine. You can also view the labels of a specific virtual machine using the name of the virtual machine:

```
phyhost$ simctl example-scenario labels uml2
uml2 : reset_ips enable_forwarding disable_forwarding
```

The other option of `simctl` (“`exec`”) allows you to manage the execution of the actions (commands) of a label. The command syntax “`simctl simname exec labelname`” executes the commands associated with the label “`labelname`” on all the virtual machines where that label is defined. For example:

```
phyhost$ simctl example-scenario exec reset_ips
Virtual machines group: uml3 uml2 uml1
OK The command has been started successfully.
OK The command has been started successfully.
OK The command has been started successfully.
Total time elapsed: 0 seconds
```

You can also run a label on a single machine with the following syntax:

```
phyhost$ simctl example-scenario exec disable_forwarding uml2
OK The command has been started successfully.
Total time elapsed: 0 seconds
```

In general, if a label is multiply defined on several virtual machines, with the previous command, the actions of a label are only executed on the specified virtual machine but not on the rest of virtual machines that have defined that label.

Finally, it is worth to mention that `simctl` executes the `start` label when any scenario is started. In addition, many

scenarios have a label called `initial`. The `initial` label is executed to load the initial configuration of the scenario. This configuration is stored in your `$HOME/.vnuml` directory. If you do not have any trouble, you will have to execute `initial` only once. But if you need to use a `forcestop` option, the initial configuration is removed.

F. Capturing Traffic

`simctl` automatically creates a `tap` interface in the `phyhost` for each virtual network. These interfaces are called `SimNet0`, `SimNet1`, etc. However, the name (`SimNet`) can be defined as desired. `Simtools` also provides a tool called `simtools-captap` to automatically start `wireshark` protocol analyzers in several `SimNetX` interfaces. The syntax is:

```
phyhost$ simtools-captap
phyhost$ simtools-captap -k to killall wiresharks
phyhost$ simtools-captap -s 7 to start capturing from SimNet0 to SimNet7
phyhost$ simtools-captap -r 7 to restart capturing from SimNet0 to SimNet7
```

V. COMPARISON WITH OTHER SYSTEMS

In our opinion, there are a couple of systems that can be used (and are actually used) to teach networks. One of these systems that can be used for similar purposes as `simtools` is the Graphical Network Simulator-3 (GNS3) [7]. GNS3 is a software emulator for networks that allows you to design complex network topologies. It allows the combination of virtual devices and real devices. You may run simulations or configure devices ranging from simple workstations to Cisco routers. It is based on `Dynamips`, `Pemu/Qemu` and `Dynagen`.

Regarding `simtools`, a problem of GNS3 for teaching is that it consumes quite high resources, mostly on scenarios with many devices. This is because its underlying virtualization technologies (`Dynamips`, `Pemu/Qemu` and `Dynagen`) consume more resources than `User Mode Linux` (the underlying technology of `simtools`). This limits the number of scenarios that you can emulate. On one hand, GNS3 is a tool clearly focused to emulate Cisco (proprietary) routers. On the other hand, `simtools` is focused only on open source and we use only Linux routers. Finally, GNS3 is not prepared to run labels, which, as explained in the next section, are a very interesting feature for teaching.

The other system that can be used for similar purposes as `simtools` is `Mininet` [8]. `Mininet` is a network emulator that runs a collection of end-hosts, switches, routers, and links on a single Linux kernel. It uses lightweight virtualization to make a single system look like a complete network, running the same kernel, system, and user code. A `Mininet` host behaves just like a real machine; inside, you can run arbitrary programs (including anything that is installed on the underlying Linux system.) The programs you run can send packets through what seems like a real Ethernet interface, with a given link speed and delay. Packets get processed by what looks like a real Ethernet switch, router, or middlebox, with a given amount of queuing.

Regarding `simtools`, `Mininet` consumes less resources because its underlying virtualization techniques (kernel namespaces and cgroups) are faster than paravirtualization (which is the technique used by UML). However, for teaching purposes, `Mininet` is a less flexible platform because `Mininet` uses the kernel of the physical host for all virtual hosts; this means

that you cannot run software that depends on other kernel versions. For example, when teaching congestion control in TCP, it is difficult to do it with newer kernels that have plenty of improvements. To teach TCP, it is much easier to use older kernels tunned, for example, with the classical `reno` or new `reno` controls. This is only an example, but in general, teaching certain parts of networking with an uncontrolled kernel is a bad idea since we can get unpredictable results. An additional problem for teaching generic networking scenarios with `Mininet` is that, by default, all `Mininet` hosts share the host file system and PID space. This means that students should be careful if they are running different daemons that require the same configuration files in `/etc` or when you have to kill a process from a certain `Mininet` host.

VI. TEACHING EXPERIENCES

In this section, we would like to mention some of our experiences when using `simtools` to teach networking.

A first issue is that students have some initial difficulties when using a command-line terminal or console. However, in our subjects we encourage our students to use this interface because the command-line is the most flexible interface. Commands can be programmed, you can build scripts and schedule them for later execution. Also observing the system logs is much natural from the command-line than from a graphical interface. In addition, the command-line is the way of configuring most network devices. We should say, that at the end, our students get used to this interface and they appreciate it and enjoy using it.

Regarding the installation and usage of `simtools`, we have been able to make it work with any system that our students had, even in quite resource limited devices like “nettops”.

The wonderful fact of having a complete network scenario emulated in a single machine thanks to virtualization is that we can also enhance the teaching experience for our students. In this context, `simtools` allows us to send homework to students that previously, only would have been possible to do at one of our laboratories. We can also support our theoretical lessons with practical examples more easily with practical real tests performed during the class with `simtools`. Furthermore, we have started to record some videos so that our students can later reproduce our configurations and tests over the emulated scenarios.

Possibilities with virtualization are numerous. For example, using the label mechanism that `simtools` provides to us, we can create complex scenarios with several configurations and these configurations are set up almost instantaneously with a single command. Many times, we use this feature (labels) for putting one of our emulated networking scenarios into a certain state and then, we propose our exercises. We also use labels to configure solutions to our practical exercises.

The authors of this paper have been using `simtools` to teach networks in five subjects at UPC during approximately the last five years. These subjects belong to two schools, they are in the teaching plans of three different degrees and a master degree. Next, we provide a little description about each subject and the topics that have been teach with `simtools`:

- Fonaments de Xarxes Telematiques (FXT). This is a subject of the “grau d’enginyeria audiovisual” at the Escola d’Enginyeria de Terrassa (EET). This subject is

the first subject about networking that these students course. We mainly teach Ethernet switching and IP routing with simtools. We have observed that students are more motivated for networks than previously and that they really think that learning networks with simtools is much easier and natural. Also the results in their marks are better since we use simtools and we feel that we can rise a little the level of the class.

- Transport Control i Gestio a Internet (TCGI). This is a subject of the "grau d'enginyeria telematica" at the Escola Tecnica Superior d'Enginyeria de Telecomunicacions de Barcelona (ETSETB). This is a rather specific subject of networking since it is in the degree of telematics. Here the level of exigency is high and the scenarios complex. Here we teach many topics of networks like advanced switching with VLANs, dynamic routing, dynamic webs, DNS, DHCP, network applications, tunnels, multicast, IPv6, and so on. For this subject, we think that a virtualization tool like simtools is essential.
- Serveis i Aplicacions Telematiques (SAT). This is a subject of the "grau d'enginyeria audiovisual" at the Escola d'Enginyeria de Terrassa (EET). This subject is the second subject about networking that these students course. Thanks to the knowledge acquired in FXT, students keep using and learning networks with simtools. In this case, we cover topics like DNS, DHCP, WEB, client/server programming etc.
- Comunicacions Multimedia (CM). This is a subject of the "grau d'enginyeria audiovisual" at the Escola d'Enginyeria de Terrassa (EET) and also a subject of the "grau d'enginyeria audiovisual" at the Escola Tecnica Superior d'Enginyeria de Telecomunicacions de Barcelona (ETSETB). In this subject we teach multimedia over Internet communications. For instance, multimedia over TCP or UDP, RTP, video streaming, Voice over IP (VoIP) and so on. For this subject, we also think that a virtualization tool like simtools is essential.
- Overlay Networks (OVNET). This is a subject of the MASTER'S DEGREE IN TELECOMMUNICATIONS ENGINEERING at the Escola Tecnica Superior d'Enginyeria de Telecomunicacions de Barcelona (ETSETB). In this subject we teach tunnels, peer to peer networks and many other overlay networks. For this subject, we also think that a virtualization tool like simtools is essential.

In summary, for us, virtualization in general, and simtools in particular, have been a revolution in the way we teach networks and how our students learn these abstract topics. Now, they can reach at great networking know how without investing in expensive equipment at their home and with a lot of supporting material like videos or solved practices.

VII. CONCLUSIONS

In this paper we presented simtools and how to install and use the tool. We compared simtools with other systems and briefly described our teaching experience with it. Simtools is successfully used by some subjects related to networking at the networking department of the Universitat Politecnica de Catalunya (UPC). The tool is open source and easy to use.

We have created a repository in which simtools is packaged to be easily installed in Debian-based Linux systems.

ACKNOWLEDGMENTS

This work was supported partially by the Spanish Research Council with Project SERVET TEC2011-26452, and by Generalitat de Catalunya with Grant 2014-SGR-1504 and 2014-SGR-375 to consolidated research groups.

REFERENCIAS

- [1] F. Galan, D. Fernandez, J. Ruiz, O. Walid, and T. de Miguel, "Use of virtualization tools in computer network laboratories," in *Information Technology Based Higher Education and Training, 2004. ITHET 2004. Proceedings of the Fifth International Conference on*, May 2004, pp. 209–214.
- [2] A. Ruiz-Martinez, F. Pereniguez-Garcia, R. Marin-Lopez, P. Ruiz, and A. F. Skarmeta-Gomez, "Teaching advanced concepts in computer networks: Vnuml-um virtualization tool," *IEEE Trans. Learn. Technol.*, vol. 6, no. 1, pp. 85–96, Jan. 2013. [Online]. Available: <http://dx.doi.org/10.1109/TLT.2013.2>
- [3] R. Y. Ameen and A. Y. Hamo, "Survey of server virtualization," *CoRR*, vol. abs/1304.3557, 2013. [Online]. Available: <http://arxiv.org/abs/1304.3557>
- [4] S. Nanda and T. cker Chiueh, "A survey of virtualization technologies," *Tech. Rep.*, 2005.
- [5] J. Dike, *User Mode Linux(R) (Bruce Perens Open Source)*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2006.
- [6] J. Watson, "Virtualbox: Bits and bytes masquerading as machines," *Linux J.*, vol. 2008, no. 166, Feb. 2008. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1344209.1344210>
- [7] "Gns3 online documentation," <http://www.gns3.com/>.
- [8] "Mininet online documentation," <http://mininet.org>.

Propuesta Metodológica para Asignaturas de Telemática en Máster Profesionales Oficiales

Elsa Macías^{1,3,4}, Álvaro Suárez^{1,3,4}, J. Aurelio Santana^{1,2,4}, Domingo Marrero^{1,3,4}, Vicente E. Mena-Santana^{1,2,4}

¹ Departamento de Ingeniería Telemática
{alvaro.suarez, elsa.macias, domingo.marrero}@ulpgc.es

² Departamento de Señales y Comunicaciones
{jose.santana, vicenteefigenio.mena}@ulpgc.es

³ Grupo de Arquitectura y Concurrencia (GAC)

⁴ Grupo de Docencia de Arquitectura y Concurrencia (GDAC)

Universidad de Las Palmas de Gran Canaria
Campus Universitario de Tafira s/n
35017 Las Palmas de Gran Canaria

Resumen- La docencia de los master oficiales en Ingeniería de Telecomunicación deben preparar al egresado para desarrollar su profesión de la mejor forma posible. En la actualidad existen muchas metodologías docentes y también retos por afrontar a corto plazo para lograr una calidad de la docencia elevada según reconocen organismos internacionales. En este artículo presentamos las ideas principales de una experiencia docente que hemos desarrollado en los últimos tres años en el master oficial de Ingeniería de Telecomunicación de la Universidad de Las Palmas de Gran Canaria. Los resultados excelentes obtenidos nos invitan a presentar la metodología seguida en varias asignaturas.

Palabras Clave- Máster Profesional, telemática, metodologías docentes.

I. INTRODUCCIÓN

Los másteres oficiales universitarios están implantados en España desde hace algunos años. Aunque es pronto para evaluar su eficacia en el proceso de enseñanza-aprendizaje, lo que sí es cierto es que el denominado *Proceso de Bolonia* [1] ha planteado nuevas oportunidades y fórmulas para desarrollar dicho proceso favoreciendo la adquisición de conocimientos significativos por parte del alumnado, haciéndole principal actor de su propio aprendizaje [2]. Para ello se pueden aplicar un sinfín de metodologías de enseñanza-aprendizaje, entre las que se encuentran: aprendizaje basado en problemas, el aprendizaje orientado a proyectos, el aprendizaje cooperativo y el contrato de aprendizaje [3]. Haciendo una búsqueda básica en la Web se puede encontrar innumerables iniciativas de las universidades españolas encaminadas a favorecer la formación de su profesorado en estas metodologías. En el mes de febrero de este año [4] se publicado el informe *Informe Horizon 2015 educación superior: tendencias, retos y tecnologías importantes* del *The New Media Consortium*. En este informe se reconoce que todavía hay mucho trabajo por realizar en el ámbito de la Universidad para desplegar

eficazmente las metodologías anteriores, marcando claramente algunas líneas que aceleran la adopción de tecnologías educativas como es el aprendizaje en línea usando recursos educativos abiertos y estrategias de cambio ágiles. Además, se marcan retos a corto plazo como son: la clase invertida (aprendizaje colaborativo y dirigido por proyectos en el Aula con la supervisión del profesor) y *Bring Your Own Device (BYOD)*.

Las enseñanzas de máster profesionales en España deben estar encaminadas a conseguir que el alumno adquiera competencias que desarrollaría en su entorno laboral con garantías de éxito. Pero esto, no es independiente de los típicos problemas de motivación de los recursos humanos. Entre los factores prácticos humanos que influyen en que se pueda conseguir estos objetivos están:

- La acción docente del profesorado. Esta acción suele ser muy positiva, puesto que en general se observa un compromiso elevado por parte del profesorado [5].
- La acción del alumnado. En asignaturas de máster, la implicación de los alumnos es muy elevada. Su nivel académico y madurez es alto por lo que suele haber una implicación considerada. Es decir, su complicidad con el profesorado para alcanzar un aventajado grado de adquisición de competencias es elevado, según hemos constatado con nuestra experiencia docente tras haber impartido en todos los niveles del ámbito universitario [6].

El objetivo anterior tampoco es independiente de los factores de gestión académica en la Universidad. Entre los que influyen están:

- La Normativa interna a nivel de Universidad. Normalmente existen reglamentos que regulan internamente la impartición de la docencia. Estos pueden imponer serias limitaciones e inversión de

tiempo por parte del profesorado, como comentamos en [7].

- Los acuerdos puntuales que se tomen en los Consejos de Departamento y en las Juntas de Escuela. Por ejemplo, algo tan simple como los horarios de impartición de las asignaturas siguen teniendo una estructura rígida e inamovible anterior al Proceso de Bolonia. Con los actuales horarios (no liberalizados), las asignaturas se deben impartir de acuerdo a una planificación semanal con un número de horas prefijado de teoría, problemas en el aula y prácticas de laboratorio.
- Los acuerdos que se lleven a cabo por parte de los equipos docentes de las asignaturas. Los profesores deben ponerse de acuerdo para implantar una determinada metodología docente de forma eficiente. En el caso que los profesores no estén de acuerdo, esto no se puede llevar a cabo, volviendo a la estructura típica anterior al Proceso de Bolonia en el que cada profesor impartía la parte de la asignatura que le competía sin interaccionar lo suficiente con sus compañeros.

En este artículo proponemos una metodología docente flexible que debe contar con el acuerdo de los profesores que la imparten, la complicitad de los alumnos, las normativas docentes vigentes y los acuerdos del Consejo de Departamento (CD) y la Junta de Centro (JC). Esta metodología está alineada con retos de clase invertida y BYOD del informe de 2015 del NMC. Además explicamos como en los últimos 3 años de impartición del máster Universitario de Ingeniero de Telecomunicación, esta metodología ha dado sus frutos excelentes en la Escuela de Ingeniería de Telecomunicación y Electrónica (EITE) de la Universidad de Las Palmas de Gran Canaria (ULPGC).

El resto del artículo se estructura como sigue: en el apartado II se presenta el entorno en el que se implanta la metodología en el contexto de la ULPGC. En el apartado III se presenta la metodología novedosa que proponemos. En el apartado IV se presentan dos casos de aplicación a dos asignaturas de máster y finalmente enumeramos algunas conclusiones de la experiencia adquirida durante estos 3 últimos años.

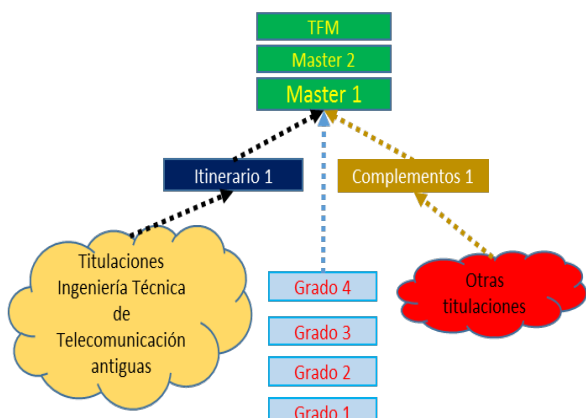


Fig. 1. Posibles caminos para realizar el Máster en la EITE-ULPGC.

II. EL ENTORNO DE GESTIÓN DE LA ULPGC

En la EITE se imparte actualmente un grado generalista de 4 cursos de Ingeniería en Tecnologías de la Telecomunicación con 4 intensificaciones: *Sistemas de Telecomunicación, Sonido e Imagen, Sistemas Electrónicos y Telemática*.

En la Fig. 1 se muestra un esquema de la estructura del máster y de los caminos posibles para acceder a él según la procedencia de los alumnos. Después de un Trabajo Fin de Grado en el cuarto curso, los alumnos pueden acceder al Máster Universitario de Ingeniero de Telecomunicación. A este Máster también pueden acceder alumnos que realicen un curso de adaptación de contenidos (denominado *itinerario de adaptación al Grado en Ingeniería en Tecnologías de la Telecomunicación*) si tienen una titulación de Ingeniero Técnico de Telecomunicación (en cualquiera de sus especialidades) previo al Proceso de Bolonia, y por otro lado pueden acceder alumnos de Ingeniería Informática, Ingeniería Industrial y otros titulados cursando unos complementos de formación. El itinerario está estructurado en un curso de dos semestres y los complementos de formación en un semestre. El Máster está estructurado en un curso completo de dos semestres (*Máster 1*), un curso de un semestre (*Máster 2*) y el Trabajo de Fin de Máster (*TFM*).

La estructura de créditos de las asignaturas asignadas al ámbito de conocimiento de Ingeniería Telemática del máster de la EITE en la ULPGC es la que se muestra en la Tabla I [8]. Como se puede observar constan de 30 horas de teoría, 13 horas de problemas en el aula, 17 horas de clases prácticas en el Laboratorio. El resto de horas corresponden a horas no presenciales y de estudio personal del alumno que deben ser planificadas por el profesor y llevadas a cabo por el alumno en cada semana.

Los proyectos docentes de estas asignaturas se llevan a cabo mediante un Proceso fuertemente reglado. Se regula: la evaluación, la planificación de la docencia, coordinación, temporalización de la docencia... Los departamentos y las escuelas, facultades o institutos universitarios (que tienen responsabilidad como centros docentes también) tienen competencias en la gestión de la docencia. En la Fig. 2 se muestra un esquema del proceso de elaboración de los proyectos docentes de las asignaturas en la ULPGC [7]. Suele durar unos dos meses aproximadamente, que consiste básicamente en lo siguiente:

1. El Vicerrectorado competente en Ordenación Académica pone en marcha el Proceso enviando a todos los centros docentes los detalles de elaboración de los proyectos: número de horas de docencia por grupo (teoría, clases práctica en el aula, clases prácticas en el laboratorio...).

Tabla I
ESTRUCTURA DE CRÉDITOS DE LAS ASIGNATURAS DEL ÁMBITO DE CONOCIMIENTO DE INGENIERÍA TELEMÁTICA EN LA EITE-ULPGC

Teoría	Problemas en el Aula	Laboratorio	No presenciales
30	13	17	92

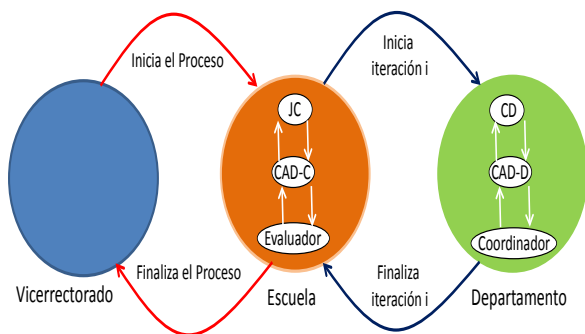


Fig. 2. Esquema de gestión docente en la ULPGC.

- Los centros, a través de sus *Comisión de Asesoramiento Docente del Centro (CAD-C)*, determinan una serie de criterios particulares que se deben cumplir para todos los proyectos docentes de las asignaturas. El *Personal de Administración (PA)* de los centros envía electrónicamente (mediante una aplicación Web) los proyectos docentes a los departamentos.
- Los departamentos elaboran los proyectos docentes después de asignar profesores a asignaturas. Los profesores coordinadores deben reunirse con los otros profesores de la asignatura para rellenarlo y someterlo al análisis de la *Comisión de Asesoramiento Docente del Departamento (CAD-D)*, las cuales verifican que el proyecto docente esté correcto (cumpla con los requisitos que marcan los distintos órganos y el VERIFICA [8]) y lo elevan al CD para que lo apruebe y el PA del departamento lo envíe electrónicamente al Centro.
- El Centro los recibe e inicia una fase de revisión asignando proyectos docentes a los miembros de la CAD-C.
- Si se detectan errores, se devuelve el proyecto docente al CD y vuelve a comenzar el Proceso en el paso 3, sino, se acaba el Proceso y el proyecto docente se somete a aprobación en la JC.

La temporalización de la docencia en estos proyectos docentes se debe adaptar al proceso anterior y en concreto a lo que acuerde la JC. Un elemento importante es que la Junta de la EITE, suele aprobar horarios de impartición de las clases de las asignaturas de Máster tal que habilitan la implantación de diferentes tipos de metodologías docentes [9]. Esto es, se permite a los profesores impartir hasta un total de 6 horas presenciales de la asignatura por semana y 2 horas de trabajo no presencial del alumno. De estas 6 horas, hasta cuatro corresponden a horas de teoría, hasta 2 horas para problemas en el Aula y hasta otras 2 para clases prácticas en el laboratorio. Cuando hay clases de problemas en el aula sólo puede haber 2 horas de teoría. En la Fig. 3 se muestra un esquema del horario de las asignaturas semanalmente con los límites establecidos. Se observa que cada semana puede tener un número flexible de horas, pudiendo desde no haber clase esa semana, hasta tener sólo dos horas presenciales y el resto de trabajo no presencial del alumno.

Teoría (2h)	Problemas (2h)	Laboratorio (2h)	Trabajo personal del alumno (2h)
(2h)	(2h)		(4h)
(2h)	(2h)		(4h)
(2h)	(2h)		(4h)
(2h)			(6h)
(2h)			(6h)
(2h)			(6h)
			(8h)

Cada semana puede tener cualquiera de estas combinaciones de clase. Incluso en una semana puede no haber ningún tipo de actividad programada

Fig. 3. Estructura del horario de clases de una asignatura en una semana.

Un elemento relevante es que los alumnos del máster no realizan prácticas en empresas.

En este marco, pensamos que se pueden implantar metodologías docentes como la que presentamos en el siguiente apartado.

III. METODOLOGÍA PROPUESTA

El máster Universitario de Ingeniero de Telecomunicación es un habilitador para desarrollar los conocimientos adquiridos en la Empresa principalmente (aunque no se puede olvidar que este Máster también habilita a los egresados a realizar un doctorado afín). Por este motivo, las asignaturas de telemática deberían formar a los alumnos en redes y servicios telemáticos que actualmente se empleen en la Empresa (sin olvidar nuevos puntos de vista relevantes en investigación que les permitan iniciarse en ella en algunos de estos aspectos concretos). Por este motivo pensamos que la impartición de la docencia en estas asignaturas debe potenciar, principalmente, la adquisición de competencias prácticas que se puedan desempeñar en el Mundo Laboral.

A. Impartición de contenidos teóricos

De las 30 horas de teoría de las asignaturas, se deben impartir 2 horas de seminario y otras 2 horas de tutoría en el Aula. Con lo cual quedan 26 horas de teoría para impartir los contenidos teóricos.

En las 2 horas de seminario se propone que un experto universitario de una empresa de redes y servicios telemáticos que desarrolle su trabajo en Gran Canaria, y a ser posible que sea egresado de la EITE, pueda impartir una charla estructurada en dos partes:

- Un *story telling* (que es una técnica antigua pero muy eficaz [10–12]) sobre su desempeño particular en su empresa explicando su experiencia concreta y cuáles son sus vivencias particulares atendiendo a las dificultades que ha tenido en desarrollar las

competencias adquiridas durante su etapa de universitario. El objetivo de esta parte es poner al alumno en contacto con la realidad y motivarle a seguir una línea de aprendizaje conducente a obtener las competencias de la mejor manera para que afiance sus conocimientos antes de desarrollarlos en la Empresa. Esto se justifica, porque en los últimos años se detecta una tendencia masiva a que el estudiante intente pasar por el Aula sin plantearse que la adquisición de conocimientos significativos es lo que le aseguraría en el futuro un desempeño óptimo de su profesión. Ellos, simplemente tienden al “mínimo esfuerzo” y cumplir con los mínimos para aprobar la asignatura.

- Una discusión sobre el estado actual del Mercado de trabajo de la Ingeniería Telemática en su entorno empresarial. El objetivo es poner en contacto al alumno con este Mercado y de la importancia de adquirir competencias significativas para que supere las barreras de entrada a dicho Mercado de la mejor forma posible. Contando con una visión privilegiada de los expertos, se logra que el alumno tome consciencia de que su entrenamiento en la Universidad no tiene tanta complejidad como la que tendría que asumir en su futuro profesional. Con ella se les ayuda a comprender que una preparación apropiada ahora les ahorraría muchos inconvenientes después. Para ello es muy útil discutir con el experto las típicas encuestas en las que se publican los perfiles más apropiados para ser desarrollados dentro de la Ingeniería telemática (previamente se les invita a que busquen datos, como por ejemplo los que suelen publicarse en [13]), en Gran Canaria, el territorio peninsular español y en cualquier otro país europeo o de otro lugar del Mundo. Estas discusiones siempre cuentan con la participación muy activa del alumno.

De las 26 horas de teoría restantes se propone dedicar un máximo del 25% al desarrollo de un proyecto de curso de baja complejidad. Este proyecto debe estar relacionado con alguna empresa del entorno en la que se desarrollen productos telemáticos. Para ello, el profesor cuenta con una base de datos de empresas que debe estar en la Bolsa de trabajo de la Universidad. Por tanto se podrían dedicar un máximo de 6,5 horas al desarrollo de este proyecto de curso. Como se cuenta entre 2 y 4 horas semanales para la teoría, entonces se puede impartir entre 7 y 13 semanas dedicándole aproximadamente media hora semanal a este proyecto. Este tiempo está destinado a establecer claramente cuál es la tarea a realizar, la supervisión semanal de las dudas que hayan ido teniendo los alumnos y la puesta en común de problemas para que entre todos los alumnos y el profesor se puedan ir resolviendo continuamente. Sólo se tiene en cuenta aspectos relacionados con los conceptos teóricos de la asignatura aplicados a ese proyecto concreto, para no desvirtuar la estructura de créditos que marca el VERIFICA.

El resultado de este trabajo tiene su reflejo en la evaluación final de la asignatura. Uno de los problemas que suelen surgir es que los alumnos se muestran reticentes a ir desarrollando el trabajo semanalmente alegando que no tienen tiempo disponible. Para resolver este problema se

invita a una parte de los alumnos a que en las clases teóricas dedicadas a esto empleen la mayor cantidad de tiempo discutiendo información con sus compañeros mientras otros diseñan la estrategia de trabajo y otro grupo va analizando las empresas. Esta estrategia tiene la ventaja que permite que entre todos trabajen en grupo y aprendan a manejar sus agendas a lo largo del curso (clases invertidas dirigidas a proyectos). Además en estas clases los alumnos implantan BYOD trabajando en línea directamente sobre La Nube para agilizar la adquisición de tareas a realizar y reflejar en tiempo real el trabajo colaborativo que realizan. La ventaja de esta técnica es que se reduce el típico *stress* del alumno que acumula tareas sin saber bien qué debe hacer porque no están claras la relación de su trabajo con el de sus compañeros. Además, en cualquier momento puede hacer copia de todo el sistema de información en su propio equipo para evitar perder información y poder trabajar fuera de línea, por ejemplo mientras está relajado en la cafetería u otro lugar en el que no tiene disponible el acceso al sistema de información. Los alumnos prefieren siempre utilizar el sistema de información directamente.

En las 2 horas de tutoría se ponen en común las tareas desarrolladas por los alumnos cooperativamente y se orienta sobre la evaluación de los resultados del aprendizaje, para facilitar que se supere la asignatura asegurando que los conocimientos adquiridos han cimentado significativamente.

B. Impartición de contenidos de problemas en el Aula

De las 13 horas de problemas en el Aula, se dedican un máximo del 60% a la discusión de aspectos sobre el proyecto de curso y el resto a la preparación de las clases prácticas de laboratorio. Esto es, unas 7 horas se dedican al proyecto de curso y 6 a preparar una introducción a las clases prácticas de laboratorio. Esto tiene sentido porque dentro de estas horas, el VERIFICA ordena que el alumno desarrolle al menos 2 horas de trabajo individual presencial para el desarrollo de trabajos y 1 hora de trabajo en grupo. Con lo cual realmente nos restan 4 horas para ir discutiendo el desarrollo del proyecto de curso. En estas 4 horas se discuten aspectos concretos de implantación del proyecto concreto para no desvirtuar la estructura de créditos del VERIFICA. Esto es, mientras en las horas de teoría se discuten aspectos del proyecto relacionados con los conceptos teóricos de la asignatura, en estas horas se discuten problemas concretos de posible implantación.

C. Impartición de contenidos prácticos de laboratorio

Las clases de laboratorio deben estar estructuradas en clases en las que el alumno desarrolla las tareas prácticas que se indiquen en el proyecto docente de la asignatura y otras de evaluación (según se especifica en el VERIFICA). Por ello, por cada tarea se recomienda dejar una hora para llevar a cabo la evaluación. Dicha evaluación continua consiste en contestar a un cuestionario tipo test sobre aspectos relevantes de la tarea práctica y la forma en cómo se ha resuelto individualmente. Esto se ve complementado por el desarrollo en grupo de las memorias de las prácticas utilizando sistemas de compartición y edición de recursos como puede ser *Google drive* u otros de la propia Universidad. El objetivo de que estas memorias se hagan en grupo es que el alumno se acostumbre a registrar sus resultados cooperando con

compañeros, ya que entendemos que en la empresa es normal encontrarse con este tipo de tareas recurrentemente.

Normalmente, los alumnos van descubriendo, poco a poco, aspectos relacionados de la teoría, problemas y prácticas de laboratorio son el proyecto que desarrollan, con lo que poco a poco se van convirtiendo en actores de su propio aprendizaje significativo.

D. La evaluación de los resultados de aprendizaje

Tal como indicamos en [14] los alumnos deben trabajar cooperativamente desarrollando contenidos teóricos de la asignatura a partir de unos contenidos iniciales que le suministra el profesor, por ejemplo utilizando *Google Drive*. Cada uno de ellos debe tener una tarea concreta a desarrollar que se suele repartir entre ellos mismos sin intervención del profesor. La participación activa en esta tarea tiene su reflejo en la evaluación de la asignatura.

Por otro lado, el desarrollo continuo del proyecto de curso en las clases de teoría tiene su reflejo mediante la técnica de listas de participación y registro personalizado cualitativo de las respuestas a preguntas en el Aula. Con esto se logra que los alumnos participen activamente.

Las clases de laboratorio tienen también un peso en la evaluación: por un lado se verifica la participación de cada alumno en la elaboración de la memoria de las tareas prácticas (utilizando funciones de registros históricos de *Google Drive*) y por otro la nota particular que obtenga de la realización del test.

Con todo lo anterior disponemos de información suficiente para poder averiguar si el alumno ha ido completando los resultados de aprendizaje típicos que se indican en el VERIFICA.

Ahora nos resta observar si ha aprendido a desarrollar un proyecto mínimo concreto y el grado de desarrollo alcanzado. Para ello se propone un método de evaluación selectivo en el que el mejor proyecto alcance la mejor nota y el resto se desvíe de ésta según el grado de desarrollo relativo alcanzado. El objetivo de esto es concienciar al alumno de la concurrencia competitiva que tendría que afrontar en el desarrollo de su profesión futura.

Un aspecto importante a evaluar en este proyecto es la capacidad del alumno para incorporar nuevos procedimientos desarrollados por él mismo que pongan de manifiesto su capacidad de desarrollar futuras tareas de investigación.

Todos estos parámetros de evaluación deben ser conocidos por todos los alumnos desde el primer día de clase. Con este compromiso de transparencia desde el primer día se logra que el alumno conozca perfectamente qué se espera de él y se motive a dar lo máximo para lograr destacar. También se consigue liberar la típica ansiedad de no conocer exactamente cuál es la forma que tiene el profesor para hacer evaluaciones. Por eso se les invita a que desde la segunda semana entreguen tareas que serán supervisadas por el profesor. Éste debe especificar con todo lujo de detalles los valores de los parámetros de evaluación justificando el valor de cada uno de ellos. De esta forma, el alumno se acostumbra a elaborar su proyecto justificando sus decisiones, y no simplemente realizándolo para cumplir y obtener la mínima nota que le permita aprobar.

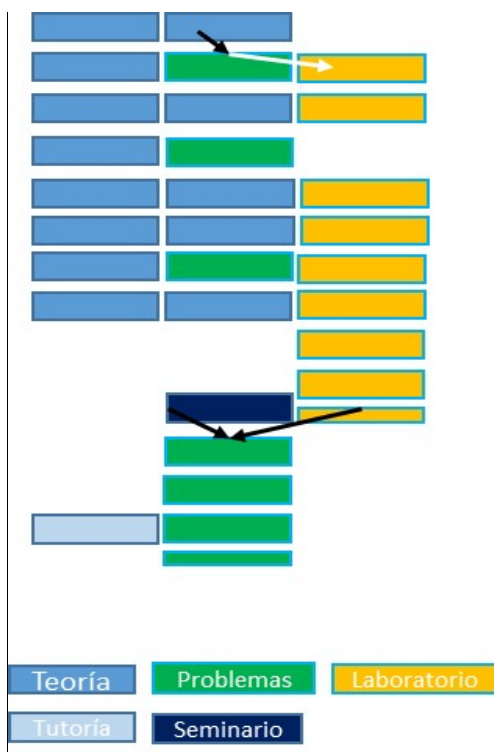


Fig. 4. Ejemplo de planificación semanal para la parte presencial de la asignatura.

E. Conjuntando todas las acciones

Una vez relatada la forma en cómo se organiza el proyecto docente de la asignatura, en este apartado presentamos una posible planificación temporal de la docencia relacionando la docencia teórica, problemas en el Aula, clases prácticas de laboratorio y otros detalles.

En la Fig. 4 se muestra un esquema que relaciona todos los elementos anteriores. Los bloques representan 2 horas (sólo hay un bloque de 1 hora en problemas en el Aula y clases de laboratorio mostrado con menor estrechez). La secuenciación en vertical corresponde a una posible planificación temporal por semanas. Las clases de problemas en el Aula intercaladas con las horas de teoría son las que se utilizan para introducir las tareas prácticas que se desarrollan en las clases prácticas de laboratorio. Se observa que existe una dependencia entre las clases de teoría y estas clases de problemas en el Aula. La clase de seminario se puede impartir después de acabar las clases de prácticas de laboratorio o un poco antes, puesto que de esa forma se puede entender mejor los planteamientos que el experto de la Empresa explique a los alumnos. Una vez acabadas las prácticas de laboratorio (o un poco antes) se inician las clases de problemas en el Aula en las que se desarrolla el proyecto de curso. Con las dos horas de tutoría que se imparten una semana antes de acabar el curso se pretende aclarar todas las dudas de teoría, y sobre todo del proyecto de curso.

IV. CASOS DE APLICACIÓN

Los tipos de proyectos de curso de los apartados anteriores están estrechamente relacionados con la asignatura

concreta en la que se desarrolle. Por este motivo en este apartado se relata la experiencia organizando la metodología docente anterior en dos asignaturas de Telemática en el Máster de la EITE.

En la Tabla II se muestran todas las asignaturas de telemática en el Máster.

Empezamos por la asignatura *Tecnologías de Internet de Nueva generación*.

Los contenidos teóricos que se imparten en esta asignatura se muestran en la Tabla III.

La idea básica es explicar los protocolos y servicios multimedia más importantes a día de hoy, los servicios Web y su implantación en una Empresa concreta. Estos servicios deben ser gestionados y asegurados mediante técnicas de gestión de la seguridad. Por tanto, uno de los resultados del aprendizaje más importantes es que los alumnos conozcan adecuadamente la implantación de nuevos servicios de Internet en una Empresa concreta.

Tabla II
ASIGNATURAS DE TELEMÁTICA EN EL MÁSTER DE LA EITE

<p><i>Tecnologías de Internet de Nueva generación</i> <i>Integración de Servicios Telemáticos en Redes Heterogéneas</i> Administración y Gestión de Redes Telemáticas Gestión de Proyectos Avanzados en Ingeniería de Telecomunicación* Ingeniería de Aplicaciones Móviles</p>
<p>(*) Compartida con otros ámbitos de conocimiento</p>

Tabla III
CONTENIDOS TEÓRICOS DE LA ASIGNATURA TECNOLOGÍAS DE INTERNET DE NUEVA GENERACIÓN

Tema 1	Tecnologías de Internet y de su acceso	4h
Tema 2	Protocolos de nueva generación de Internet para servicios telemáticos multimedia de tiempo real	6h
Tema 3	Servicios de Internet de nueva generación	8h
Tema 4	Introducción al diseño de aplicaciones móviles para Internet	4h
Tema 5	Herramientas y estrategias de gestión de la seguridad	4h

Tabla IV
CONTENIDOS DE PROBLEMAS EN EL AULA DE LA ASIGNATURA TECNOLOGÍAS DE INTERNET DE NUEVA GENERACIÓN

Tema 1	Ayuda a la comprensión de la práctica de VoIP	2h
Tema 2	Ayuda a la comprensión de la práctica de Portal Web	2h
Tema 3	Ayuda a la comprensión de la práctica de diseño de una aplicación móvil	2h
Tema 4	Diseño de un servicio multimedia para una empresa	7h

Tabla V
CONTENIDOS DE PRÁCTICAS DE LABORATORIO DE LA ASIGNATURA TECNOLOGÍAS DE INTERNET DE NUEVA GENERACIÓN

Tema 1	Observación de tráfico de protocolos de gestión de red	2h
Tema 2	Instalación y manipulación detallada de un servicio de VoIP	5h
Tema 3	Manejo básico de herramientas de diseño de portales Web	3h
Tema 4	Implementación de una aplicación móvil sencilla	4h
Tema 5	Manejo de herramientas de seguridad	2h

Los contenidos de problemas en el Aula que se imparten en esta asignatura se muestran en la Tabla IV.

Se puede observar que hay tres sesiones de 2 horas cada una dedicada a preparar el desarrollo posterior de prácticas de laboratorio, y otro problema que se desarrolla en 7 horas. Este trabajo está recogido en el VERIFICA como trabajos teóricos y prácticos que se debe llevar a cabo de forma individual y hacer una sesión de 2 horas de trabajo en grupo para poner en común la integración de todos los trabajos en un trabajo de mayor envergadura.

Los contenidos de prácticas de laboratorio que se imparten en esta asignatura se muestran en la Tabla V.

Cuando el alumno ya ha cursado los contenidos teóricos y realizado la mayoría de los contenidos prácticos, y además ha ido siendo orientado, a medida que se imparten las sesiones de media hora semanal dentro del bloque de horas de teoría, está en condiciones de afrontar el desarrollo (durante casi un mes) de un trabajo en el que desarrolle todos los contenidos explicados previamente.

Al finalizar el tema 1 de teoría se le presenta un escenario que consiste en un hotel el cual debe disponer de una red de acceso a un operador de telecomunicación tradicional. Se invita al alumno a que en grupo investigue sobre los hoteles de Gran Canaria y los tipos de redes de acceso que disponen. Después del tema 3 se les invita a que identifiquen las necesidades de servicios de *Voz sobre Internet Protocol (VoIP)* que podrían ser implantados en el hotel que haya escogido el alumno. Cuando finaliza el tema 4 se sugiere que se diseñe una aplicación móvil para turistas de esos hoteles y que resuelva un problema concreto: por ejemplo, eliminar la necesidad de implantar teléfonos tradicionales en cada habitación del hotel y usar VoIP móvil en su lugar. Al finalizar el tema 5 se anima a que estudien cómo implantarían la seguridad de los servicios en el hotel. Cuando finalizan las prácticas, los alumnos tienen un conocimiento práctico de los problemas con los que se encuentran en el laboratorio, por lo que al comenzar el desarrollo del proyecto de curso, ya están en condiciones de elaborar un proyecto en el que recoger aspectos teóricos y prácticos de los servicios a implantar en el hotel.

Durante las restantes 7 horas, se llevan a cabo sesiones de trabajo individual y acción tutorial del profesor en el Aula para resolver todos aquellos problemas que van surgiendo.

Esta metodología se ha llevado a cabo en los 3 últimos cursos. Además, cada año se va reutilizando el material elaborado en el curso anterior. En el último curso, los alumnos, cooperativamente, han desarrollado una memoria en la que se recoge un pliego de condiciones muy elaborado sobre la implantación de un sistema de VoIP seguro en varios hoteles existentes en Gran Canaria. Además, se implantó un lugar Web para los turistas en el que se puede reservar habitaciones y realizar llamadas de VoIP. Por último se desarrolló una aplicación móvil de VoIP para los turistas del hotel.

La experiencia es excelente puesto que la nota media de los alumnos fue notable 8. Además de que en la encuesta final manifestaron haber aprendido contenidos que pueden implantar en una Empresa real. De esta manera, el alumno pierde el típico miedo a afrontar su Vida profesional por miedo a no saber cómo desarrollar sus conocimientos.

La otra asignatura en la que hemos desarrollado esta metodología de enseñanza-aprendizaje es *Integración de servicios telemáticos en redes heterogéneas*.

Tabla VI
CONTENIDOS TEÓRICOS DE LA ASIGNATURA INTEGRACIÓN DE SERVICIOS
TELEMÁTICOS EN REDES HETEROGÉNEAS

Tema 1	Introducción a la integración de servicios telemáticos en redes heterogéneas	4h
Tema 2	Integración de servicios en redes de acceso a Internet	6h
Tema 3	Integración de servicios en redes troncales: Internet	8h
Tema 4	La Web como plataforma de integración de servicios	4h
Tema 5	Integración de servicios basados en sensores	4h

Tabla VII
CONTENIDOS DE PROBLEMAS EN EL AULA DE LA ASIGNATURA INTEGRACIÓN
DE SERVICIOS TELEMÁTICOS EN REDES HETEROGÉNEAS

Tema 1	Ayuda a la comprensión de la práctica de VoIP	2h
Tema 2	Ayuda a la comprensión de la práctica de IPTv	2h
Tema 3	Ayuda a la comprensión de la práctica de integración de servicios mediante la Web	2h
Tema 4	Diseño de una solución de integración de servicios para Empresa	7h

Tabla VIII
CONTENIDOS DE PRÁCTICAS DE LABORATORIO DE LA ASIGNATURA
INTEGRACIÓN DE SERVICIOS TELEMÁTICOS EN REDES HETEROGÉNEAS

Tema 1	Observación de tráfico de VoIP sobre una Red Privada Virtual	6h
Tema 2	Instalación y manipulación detallada de un servicio de IPTv en una red heterogénea	5h
Tema 3	Uso de interfaces Web para proveer el acceso a un servicio telemático integrado en redes heterogéneas	5h

Los contenidos teóricos que se imparten en esta asignatura se muestran en la Tabla VI.

La idea básica es presentar la integración de servicios en las redes de acceso y en las redes troncales haciendo énfasis en Internet. Se analiza la Web como plataforma universal de implantación de servicios telemáticos y por último se hace un análisis de los nuevos servicios telemáticos basados en sensores. A lo largo de los temas teóricos se van poniendo ejemplos de nuevos servicios típicos como son Smart Grid, Smart city, Smart home... En detalle se estudian dos tipos de servicios básicos telemáticos: la VoIP y la *Televisión sobre IP (IPTv)*. En concreto se analiza cuáles son las redes de acceso que proveen un mejor servicio para ellos y la problemática de la integración de estos servicios en redes heterogéneas. Se aplica la teoría de tráfico para dimensionar adecuadamente estos servicios y se analiza la planificación de las redes para que los soporten adecuadamente. Y en cada caso se va analizando los problemas típicos de esos servicios: su gestión, la seguridad y la legislación vigente en la actualidad.

Los contenidos de problemas en el Aula que se imparten en esta asignatura se muestran en la Tabla VII.

Se puede observar que existen sesiones de clases de problemas para preparar la práctica de VoIP sobre *Virtual Private Network* con terminales móviles, la implantación de un sistema de televisión híbrido usando la tecnología de *ChromeCast* [15], y por último el diseño de una práctica de integración de los servicios anteriores a través de interfaces

Web en el móvil. Con la realización de estos problemas se puede afrontar con éxito las prácticas en el laboratorio.

Los contenidos de prácticas de laboratorio que se imparten en esta asignatura se muestran en la Tabla VIII.

Al finalizar las prácticas, el alumno conoce perfectamente detalles sobre las distribuciones de tráfico para llamadas de VoIP variando el número de usuarios del servicio. Esto es interesante porque se puede observar que estas distribuciones de tráfico no suelen coincidir con las que se estudian teóricamente en la asignatura *Administración y gestión de redes telemáticas*. Y comprenden que en la práctica deben dimensionar y planificar las redes haciendo experimentación *in situ*. Cuando instalan un servicio local de IPTv observan las dificultades de transportar el tráfico en redes que usan segmentos cableados a alta velocidad (*Gigabit Ethernet*) y luego pasarlo por canales inalámbricos (*Wireless Fidelity*). Comprenden que deben proveer servidores o pasarelas para adaptar el tráfico entre redes. Por último cuando realizan la práctica de interfaces Web para los servicios de VoIP e IPTV observan la facilidad con la que usando *Web Real Time Communications (WebRTC)* [16] se puede implantar una interfaz de este tipo, sin perder de vista la complejidad del transporte de los datos en las redes heterogéneas y su dimensionado de tráfico real.

Afrontar el desarrollo del proyecto de curso, una vez el alumno ha sido orientado en las sesiones de teoría, y ha realizado los problemas en el Aula y las prácticas, es un reto que les permite adquirir conocimientos significativos sobre un caso real.

Para ello, una vez finalizado el tema de teoría 2 se les invita a que dimensionen la red de acceso y la red interna del hotel que hubieran elegido en la asignatura *Tecnologías de Internet de Nueva Generación* (que han debido cursar en el primer semestre). Terminado el tema 3 se les invita a que estudien los problemas que surgen cuando se conectan varios hoteles entre ellos para confederar diferentes servicios entre ellos y que analicen la problemática que surge cuando los turistas quieren usar VoIP e IPTv. Para ello tienen que aplicar sus conocimientos de modelado de tráfico que previamente han estudiado en otra asignatura. Al finalizar el tema 4 se les invita a que analicen cómo podrían usar la WebRTC para modernizar las interfaces de VoIP e IPTv en los hoteles, eliminando de esta forma equipos antiguos y proveyendo una simple pantalla táctil para implantar estos servicios, o bien el teléfono móvil de los turistas (y clientes en general). Por último al finalizar el tema 5 se les invita a que analicen cómo se puede emplear los sensores para mejorar los servicios telemáticos de los hoteles actuales.

Durante las restantes 7 horas, se llevan a cabo sesiones de trabajo individual y acción tutorial del profesor en el Aula para resolver todos aquellos problemas que van surgiendo.

Esta metodología se ha llevado a cabo en los 2 últimos cursos. Además, cada año se va reutilizando el material elaborado en el curso anterior. En el último curso, los alumnos, cooperativamente, han desarrollado una memoria en la que se recoge un pliego de condiciones muy elaborado sobre la implantación de las redes y los servicios de VoIP e IPTv en distintos hoteles reales de Gran Canaria. Se modeló el tráfico utilizando teoría de colas y se confrontó con el tráfico real medido las prácticas de laboratorio observando las diferencias. Se documentó el estudio de la integración de servicios de distintos hoteles haciendo uso de redes

inalámbricas con distintas tecnologías y también del acceso a Internet.

La experiencia es excelente puesto que la nota media de los alumnos fue notable 8,5. Igualmente manifestaron, en la encuesta final haber tomado contacto real con la verdadera problemática que supone implantar un servicio sobre una Empresa real y cuidar todos los detalles como la legislación vigente, la planificación y dimensionado eficiente de los servicios y las redes...

V. CONCLUSIONES

Las asignaturas de Telemática en el Máster Universitario de Ingeniero de Telecomunicación en la EITE que nosotros hemos implantado en los últimos años han tenido un carácter profundamente aplicado a la Empresa. El objetivo ha sido que el alumno tome consciencia de cuáles serían sus atribuciones cuando desarrolle su profesión. Para ello hemos empleado una metodología de enseñanza-aprendizaje que se basa en el desarrollo de un proyecto de curso (recogido en el VERIFICA como trabajo teórico-práctico), en el que demuestre que es capaz de enfrentarse a la instalación de servicios telemáticos y de dimensionar adecuadamente las redes.

La experiencia ha sido muy gratificante y provechosa, toda vez que en los tres últimos cursos el éxito académico ha sido muy elevado. Además, el complemento que supone traer a un experto universitario en telecomunicación que desarrolla su trabajo en la Empresa supone un aliciente muy elevado que les ayuda a desarrollar su proyecto de curso.

Una cuestión pendiente es pensar en que si en todas las asignaturas del Máster se implantara la misma estrategia, entonces se podría cargar demasiado las primeras semanas con contenidos de clases presenciales teórico y de problemas en el aula. Sin embargo, no todas las asignaturas del Máster se prestan a implantar esta metodología: tienen contenidos matemáticos muy profundos y se desarrollan problemas típicos de teoría de la comunicación y electrónica. Por ello, se entiende que es posible llevar a cabo esta planificación temporal de la docencia y vale la pena dado la orientación práctica de las asignaturas de telemática y su directa proyección en la Empresa.

Una posible línea de trabajo, en coordinación con la Directiva de la EITE, sería organizar una sesión final, en la que se invitara a un panel de expertos que analizaran los trabajos realizados por los alumnos de cara a que pudieran ser aconsejados sobre su viabilidad económica y de negocio. Si se invitara a empresarios turísticos quizás sería una buena oportunidad para dar a conocer las competencias adquiridas por los alumnos de cara a encontrar trabajo en sus empresas.

AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación parcial de la ULPGC.

REFERENCIAS

- [1] E.K. Bob Reinalda, *The Bologna Process - Harmonizing Europe's Higher Education*, Barbara Budrich Publishers, Opladen & Bloomfield Hills, 2005.
- [2] M. del Rosario González Martín y Gonzalo Génova Fuster, "Innovación docente a la luz de Bolonia: trabajo en equipo y revisiones

- cruzadas para convertir al alumno en protagonista de su proceso de aprendizaje," *Revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, vol. 9 (1), 2008, pp. 126–141.
- [3] M. de Miguel Díaz (coordinador), *Metodologías de enseñanza y aprendizaje para el desarrollo de competencias. Orientaciones para el profesorado universitario ante el Espacio Europeo de Educación Superior*, Alianza Editorial, 2006.
- [4] T.N.M.C. (NMC), "Informe Horizon 2015 educación superior: tendencias, retos y tecnologías importantes. Disponible electrónicamente en: <http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf>," 2015.
- [5] T. Ariza, "Análisis de la satisfacción del profesorado con el espacio europeo de educación superior: propuestas de mejora para la convergencia europea," Tesis Univ. Granada. Programa Oficial de Doctorado en: Psicología.
- [6] Jorge Azorín López et al., "El proceso de enseñanza-aprendizaje en la formación interuniversitaria: una experiencia a nivel de máster universitario," *XII Jornadas de Redes de Investigación en Docencia Universitaria, Universidad de Alicante, 2014. ISBN 978-84-697-0709-8, pp. 1400-1415*, 2014.
- [7] Elsa Macías, Alvaro Suarez, J. Aurelio Santana, Domingo Marrero, V. Mena, J.M. Canino, "Propuesta de mejora de la coordinación temprana vertical y horizontal de asignaturas," *I Jornadas Iberoamericanas de Innovación Educativa en el ámbito de las TIC Las Palmas de Gran Canaria – 27-28 de Noviembre 2014, ISBN: 978-84-617-1860-3, pp. 289-298*, 2014.
- [8] "Plan de estudios del Master Universitario de Ingeniería de Telecomunicación de la Universidad de Las Palmas de Gran Canaria, Registro de Universidades, Centros y Títulos (RUCT), <https://sede.educacion.gob.es/cid/75415411590421941583942.pdf>," 2015.
- [9] "Horario del Master de Ingeniería de Telecomunicación de la Universidad de Las Palmas de Gran Canaria. <http://www.eite.ulpgc.es/index.php/horarios/master>," 2015.
- [10] M.L. Gomez, A.B. Walker, and M.L. Page, "Personal experience as a guide to teaching," *Teaching and Teacher Education*, vol. 16, 2000, pp. 731–747.
- [11] F.E. Luwisch, "Understanding what goes on in the heart and the mind: learning about diversity and co-existence through storytelling," *Teaching and Teacher Education*, vol. 17, 2001, pp. 133–146.
- [12] M.J. Shank, "Teacher storytelling: A means for creating and learning within a collaborative space," *Teaching and Teacher Education*, vol. 22, 2006, pp. 711–721.
- [13] "Expansion.com (2014). Los perfiles mas demandados en 2015... Y los países que los reclaman. Available at: <http://www.expansion.com/2014/12/12/emprendedores-empleo/mercado-laboral/1418412177.html> (Accessed: 11 May 2015)," 2015.
- [14] Á.Suárez, Elsa Macías, "Organización de docencia cooperativa usando Google drive," *Jornadas de Ingeniería Telemática (JITEL 2013)*, ISBN: 978-84-616-5597-7, 2013.
- [15] F. MacNeill, "Approaching apps for learning, teaching and research," A. Middleton, ed., *MELSIG*, 2015, pp. 238–264.
- [16] H. Oh, S. Ahn, J. Choi, and J. Yang, "WebRTC Based Remote Collaborative Online Learning Platform," *Proceedings of the 1st Workshop on All-Web Real-Time Systems*, Bordeaux, France: ACM, 2015, pp. 9:1–9:5.

Una experiencia de “Flipped Classroom” en Ingeniería Telemática

Iván Marsá Maestre, Juan R. Velasco, Enrique de la Hoz y José Manuel Giménez-Guzmán

Departamento de Automática

Universidad de Alcalá

Edificio Politécnico, Campus Universitario, 28805 Alcalá de Henares, Madrid.

{ivan.marsa, juanramon.velasco, enrique.delahoz, josem.gimenez}uah.es

Resumen- La metodología de clase invertida o *flipped classroom* propone invertir el flujo tradicional de actividades de aprendizaje, realizando las actividades de “transmisión de conocimiento” fuera de clase (e.g. mediante lecturas o vídeos), y aprovechando las clases propiamente dichas para realizar actividades de aprendizaje activo más centradas en el alumno que en el profesor. En este artículo se describe una experiencia de *flipped classroom* que se ha realizado desde el Área de Ingeniería Telemática de la Universidad de Alcalá en el contexto de una asignatura troncal de redes en el curso 2014/2015.

Palabras Clave- *flipped classroom*, aprendizaje activo

I. INTRODUCCIÓN

La idea de introducir cambios en el proceso docente para mejorar la experiencia y rendimiento de nuestros estudiantes no es nueva ni impopular. El artículo de finales de los 80 *Seven Principles for Good Practice in Undergraduate Education* [1] tiene más de 4000 citas¹ (eso son algo más de 140 citas por año de publicación), y se utiliza como lectura recomendada en cursos sobre docencia de grado en universidades de prestigio². Sin embargo, estas “buenas prácticas” no parecen terminar de integrarse de forma generalizada en nuestra docencia de grado. La clase magistral expositiva, seguida del estudio autónomo por parte del estudiante de los contenidos vistos en clase sigue siendo la metodología predominante en nuestras asignaturas. En el extremo opuesto del espectro metodológico, surge el concepto de *flipped classroom* o clase invertida [2], en el que se plantea el trabajo autónomo previo del alumno del material de la asignatura, y la posterior realización de actividades de aprendizaje activo en el aula, aprovechando la oportunidad única de la presencialidad para la interacción profesor-alumno y entre alumnos. El debate sobre cuál de estos dos extremos metodológicos (o de sus múltiples posibles puntos intermedios) debe prevalecer en la enseñanza universitaria está lejos de estar zanjado [3], pero lo que sí parece aunar a investigadores en educación es la idea de que los argumentos al respecto se basen en evidencias más que en ideologías, costumbres o tendencias [4].

Con ese espíritu, hemos realizado dentro del Área de Ingeniería Telemática de la Universidad de Alcalá (UAH) una experiencia de *flipped classroom* en una de las asignaturas troncales sobre redes de ordenadores. Este artículo pretende describir y valorar esa experiencia. Comenzamos con un breve repaso a los antecedentes de este tipo de iniciativas (Sección II), para después pasar a describir el contexto y el desarrollo de nuestra experiencia particular (Secciones III y IV). Finalmente, se presenta un análisis de los resultados obtenidos respecto a diferentes indicadores (Sección V), y se identifican las principales conclusiones y líneas futuras de trabajo.

II. ANTECEDENTES

A. Buenas prácticas en educación de grado

Chickering y Gamson identifican siete principios fundamentales como buenas prácticas en la educación universitaria de grado [1]:

1. Contacto entre profesores y alumnos: el contacto frecuente profesor-alumno dentro y fuera del aula se presenta como favorecedor de la motivación e implicación del alumno en su propio proceso de aprendizaje. Más allá de ser meros vehículos para la transmisión del conocimiento, los profesores ejercen también, para bien o para mal, de modelos actitudinales.
2. Reciprocidad y cooperación entre estudiantes: el aprendizaje mejora cuando es un proceso colaborativo y social, mejorando no sólo la implicación y responsabilidad, sino también el entendimiento y la capacidad de pensamiento crítico.
3. Aprendizaje activo: los alumnos deben ser protagonistas de su aprendizaje, no espectadores. El aprendizaje es más efectivo cuando el alumno se implica directamente en el proceso.
4. Realimentación temprana: el aprendizaje de calidad se fundamenta en el reconocimiento de lo que se sabe y de lo que no. Los alumnos necesitan múltiples oportunidades durante el curso para reflexionar de forma crítica acerca de lo que han aprendido y de lo que les queda por aprender.
5. Gestión del tiempo: el realismo en la asignación de tareas y la ayuda al estudiante para una gestión eficaz de su tiempo son críticos para facilitar el aprendizaje.
6. Expectativas elevadas: las expectativas de rendimiento que tenemos sobre nuestros estudiantes se convierten con frecuencia en “profecías autocumplidas”, en especial

¹ Fuente: Google Scholar

² Véase, a modo de ejemplo, el curso sobre docencia universitaria que la Johns Hopkins University ofrece a través de la plataforma de cursos masivos online Coursera: (<https://www.coursera.org/course/univteaching101>).

cuando se traducen en un mayor o menor esfuerzo (o entusiasmo) por parte del docente.

7. *Diversidad en los mecanismos de aprendizaje*: el camino al aprendizaje no es único, y diferentes estrategias pueden funcionar mejor para unos estudiantes que para otros. La diversidad permite que todos los estudiantes tengan la oportunidad de aprender de la forma que les resulte más fácil, lo que facilitará que se esfuercen en aquellas que les resulten más costosas.

Pese a la popularidad del artículo de Chickering y Gamson, estas buenas prácticas no predominan en nuestro sistema de educación superior, y especialmente en ingeniería. Los alumnos rara vez aparecen por nuestras tutorías (salvo la semana antes del examen, probablemente). Nuestras mayores inquietudes acerca del trabajo colaborativo de los estudiantes son la copia (que unos alumnos copien el trabajo de otros) y el trabajo en equipo “asimétrico” (que unos miembros del grupo trabajen significativamente más que otros). A menudo toda la realimentación que reciben de nosotros nuestros alumnos son las notas de las pruebas parciales, y a menudo varias semanas después de la realización de la prueba correspondiente. No son raras las prácticas de laboratorio que absorben la totalidad del tiempo de trabajo del alumno (llegando a impactar en otras asignaturas), ni tampoco los comentarios pesimistas por parte del profesorado acerca de la capacidad o la dedicación de los alumnos. Finalmente, la clase magistral sigue siendo la estrategia metodológica predominante en nuestras aulas. No hay duda de que, casi treinta años después del artículo al que nos referimos, todavía hay mucho espacio para la mejora en esas “buenas prácticas”.

B. La metodología de “flipped classroom”

El término de “clase invertida” o *flipped classroom* alude a la idea de invertir el escenario tradicional donde tienen lugar las actividades en educación, realizando la “transmisión unidireccional de conocimiento” (tradicionalmente asociada a la clase magistral) fuera del aula, y realizando en clase actividades de asimilación y aprendizaje protagonizadas por el alumno (incluidas las que tradicionalmente se mandan para casa), pero con la supervisión y el apoyo del profesor [5]. Parte de la base de que, en general, la clase magistral es bastante unidireccional y poco interactiva, por lo que puede sustituirse con facilidad por lecturas, vídeos o discusiones *online* (e.g. empleando foros), de forma que la clase presencial se aproveche para aquellas actividades que puedan beneficiarse de la coincidencia de profesores y alumnos en un mismo espacio y tiempo. Algunos de esos beneficios son las oportunidades para el trabajo colaborativo y la realimentación inmediata, además de permitir al profesor seguir “en tiempo real” los procesos de razonamiento de sus estudiantes, frente a la perspectiva tradicional de ver únicamente el resultado de dichos procesos, cuando entregan sus tareas o se enfrentan al examen.

Existe un cuerpo de investigación considerable acerca de inversión de clases [6], y multitud de recursos online sobre el tema³. Sin embargo, los profesores seguimos siendo reacios

al cambio. En ingeniería, en especial, tendemos a escudarnos en la idea de que la ciencia y la tecnología están “menos abiertas a discusión” que otras disciplinas y en que, especialmente en asignaturas básicas, “la teoría es fundamental”. La clase invertida no tiene por qué ser la mejor manera de facilitar el aprendizaje de nuestros alumnos, pero es cierto que la clase magistral ya no parece resultar atractiva a nuestros estudiantes. Nuestras clases “de teoría” cada vez están más vacías. Y en cierto modo es comprensible. Con las posibilidades que brindan las nuevas tecnologías para el acceso global al conocimiento técnico, la enseñanza presencial universitaria debe plantearse cómo convertir su factor diferenciador (esa presencialidad) en un valor añadido para el estudiante. Buscando ese valor añadido que haga que nuestros alumnos quieran ir a clase hemos decidido hacer este experimento.

III. CONTEXTO DE LA EXPERIENCIA

La experiencia de *flipped classroom* objeto de este trabajo se plantea en el contexto de la asignatura *Arquitectura de Redes II (AR2)*, que se imparte como asignatura obligatoria en la UAH en el segundo cuatrimestre del segundo curso de las titulaciones de Grado en Ingeniería Telemática (GIT), Grado en Ingeniería de Sistemas de Telecomunicación (GIST), Grado en Ingeniería en Electrónica de Comunicaciones (GIEC) y Grado en Ingeniería de Tecnologías de Telecomunicación (GITT). Junto con la asignatura *Arquitectura de Redes I (AR1)*, que se imparte en el primer cuatrimestre del mismo año, constituyen el primer contacto de los alumnos con la Ingeniería Telemática. En estas asignaturas se realiza una panorámica de los diferentes niveles del modelo de referencia de TCP/IP, siguiendo un enfoque descendente, tal y como se describe en el libro de Kurose [7]. En concreto, la asignatura AR2 cubre los niveles de red y enlace, además de presentar los desafíos más importantes de las redes inalámbricas, la seguridad de redes y la gestión de red. Es una asignatura de 6 ECTS, que se imparte tradicionalmente con dos horas semanales de clases magistrales de grupo grande, y dos horas semanales de grupo pequeño dedicadas fundamentalmente a resolución de problemas y prácticas de laboratorio. Para más información acerca de la asignatura, puede consultarse su guía docente⁴.

La asignatura de AR2 nos pareció idónea para esta experiencia por diversos motivos:

- Se trata de una asignatura básica de segundo curso, sin requisitos de aprendizaje previos dependientes de otras asignaturas (a excepción de AR1), lo que permite aislar la contribución del cambio metodológico de posibles diferencias de conocimientos previos de los alumnos.
- Es una asignatura que lleva impartándose desde el curso académico 2010/2011, y que cuenta con abundante material didáctico en forma de transparencias y ejercicios con y sin solución, además de un texto de referencia [7] que cubre la práctica totalidad de la asignatura. Esto hace que los alumnos dispongan de recursos para la parte de aprendizaje previo fuera del aula propia del modelo de *flipped classroom* sin

³ Véase como ejemplo <http://flippedclassroom.org/>

⁴http://www.uah.es/estudios/asignaturas/descarga_fichero.asp?CodAsig=350015&CodPlan=G37&Anno=2014-15

necesidad de hacer un gran esfuerzo de creación de material.

- Es una asignatura que tiene múltiples grupos grandes y pequeños, lo que permite concentrar la experiencia en uno solo de los grupos, manteniendo el resto como grupos de control a efectos comparativos.

Dentro de la asignatura AR2, el grupo que se escogió para la experiencia fue el grupo de docencia en inglés, que presenta condiciones muy favorables para esta metodología docente, tanto por el número de alumnos (12 alumnos matriculados en el curso 2014-2015) como por las condiciones físicas del aula en la que se imparte (mesas y sillas móviles). Ambas condiciones favorecen el desarrollo interactivo de las clases y la realización de actividades en grupo, que veremos que son piezas clave de la experiencia. Finalmente, con el fin de acotar el esfuerzo de preparación de las sesiones, la experiencia se restringió al grupo grande (las clases de grupo pequeño se mantuvieron inalteradas).

IV. DESARROLLO DE LA EXPERIENCIA

A. Principios generales

De acuerdo con las “*best-practices*” de las que se hablaba en la Sección II, y con la experiencia previa en las clases de grupo grande, la metodología para la experiencia se estructuró alrededor de los siguientes principios:

- Redundancia (casi) cero: el trabajo personal del alumno durante la semana sería (principalmente) leer los apartados relevantes del libro de texto y el material de apoyo correspondiente. Como consecuencia, en las clases de grupo grande *no* se explicaría ese material. Si habría, no obstante, una breve introducción en los cinco últimos minutos de cada clase acerca de los contenidos que los alumnos tendrían que trabajar para la siguiente semana, a modo de motivación.
- Aprendizaje activo y colaborativo: las clases de grupo grande se dedicarían a *trabajar* sobre el material que los alumnos ya habían estudiado antes de venir a clase, planteando diferentes actividades, generalmente para hacer en parejas o en pequeños grupos.
- Expectativas elevadas: las actividades que se realizaran en clase irían en orden creciente de dificultad, sobre la premisa de que los alumnos ya habían trabajado el material en casa. Los objetivos de aprendizaje estarían en su mayoría en los niveles 3 y 4 de la taxonomía de Bloom [8] (aplicación y análisis), asumiendo que la comprensión ya había sido alcanzada por el alumno (o resuelta en tutorías/foros/correo electrónico antes de la clase). Del mismo modo que nuestras expectativas hacia los alumnos serían elevadas, asumimos que las de ellos hacia nosotros también lo serían. Nosotros esperábamos que ellos trabajaran duro en clase; ellos esperarían que hiciéramos que las clases merecieran la pena.
- Realimentación temprana: durante las sesiones de grupo grande, los alumnos tendrían múltiples oportunidades de contrastar su aprendizaje, de enfrentarse a preguntas y de contrastar sus respuestas con las de sus compañeros y con el punto de vista del profesor. Una vez finalizada la sesión, el alumno tendría en un plazo breve (uno o dos días) una calificación asociada a su participación en la sesión. De forma similar, al finalizar cada sesión los

Addressing, netmasks, forwarding and subnetting

❖ Could the two “extreme” machines communicate?

- A. Yes
- B. No
- C. Only left to right
- D. Only right to left

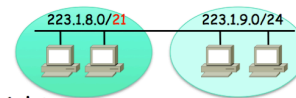


Fig. 1. Ejemplo de pregunta a la clase

alumnos tendrían la posibilidad de evaluar el desarrollo y la metodología de la sesión de forma voluntaria y anónima por medio de una encuesta *online*.

B. Actividades realizadas

Aunque, por su naturaleza interactiva y centrada en el alumno, las sesiones de *flipped classroom* son difíciles de predecir, en nuestro caso tratábamos de seguir siempre la misma estructura:

- Comenzábamos la clase con un *Icebreaker*, una breve exposición o actividad no directamente relacionada con la clase, pero que servía de introducción o motivación ilustrativa a la misma. Por ejemplo, en una clase sobre las funciones de encaminamiento y reenvío en internet se empleó un video sobre un modelo “físico” de Internet⁵.
- Listábamos en una transparencia los epígrafes de contenido que habían tenido que trabajar en casa, y dábamos la oportunidad a los alumnos de plantear cualquier pregunta que tuvieran.
- Una vez resueltas las dudas, realizábamos una serie de actividades relacionadas con los contenidos de la sesión.
- Finalmente, concluíamos con una breve presentación de los contenidos que tendrían que trabajar para la siguiente semana.

Entre las actividades que se realizaron en clase, cabe destacar las siguientes:

- Preguntas tipo “test” a la clase: la actividad más básica, sencilla y versátil, y una de las más utilizadas. El profesor plantea una pregunta a la clase mediante una transparencia y proporciona una serie de posibles respuestas. Los estudiantes tienen que escoger la respuesta correcta, primero por sí mismos, después discutiendo en pareja. Las preguntas se lanzaban en orden de dificultad creciente, a menudo incluyendo sutilezas poco obvias y yendo un poco más allá del alcance de la asignatura (un ejemplo en la Figura 1). Para que los alumnos respondieran a las preguntas, empleamos la aplicación *Socrative*⁶, que permite a los alumnos contestar desde sus teléfonos móviles, y al profesor ver la estadística de respuestas.
- Resolución de problemas por parejas/grupos: de acuerdo con la metodología de *Problem-Based Learning (PBL)* [9], se plantearon problemas en clase que los

⁵ <https://www.youtube.com/watch?v=B4auh-YFFfE>

⁶ <http://www.socrative.com/>

DHCP renewal



What if the client wants to renew but the DHCP server is down?

- A. It will keep its current IP
- B. It will try again until it succeeds
- C. It will search for other DHCP servers
- D. It will get itself a new address of the form 169.254.X.X



Fig. 2. Ejemplo de pregunta enviada por un estudiante

alumnos debían resolver por parejas o por grupos. De nuevo, el grado de dificultad de los problemas a lo largo de la sesión era creciente, partiendo de pequeños problemas de rápida resolución y llegando a problemas más parecidos a los que podrían encontrarse en un examen. Durante la realización de los problemas, el profesor iba supervisando a las diferentes parejas/grupos, tratando de facilitar su aprendizaje. En los problemas de resolución más larga, el profesor alternaba preguntas relacionadas (de nuevo empleando *Socratic*) que permitían clarificar conceptos o solucionar bloqueos.

- **Dramatizaciones:** para aquellos conceptos más críticos o complejos, se diseñaron actividades en las que los alumnos podían ser partícipes directos de una recreación del proceso que se analizaba en la sesión. Por ejemplo, se realizó una dramatización del algoritmo de encaminamiento de vector-distancia, en la que cada pareja de alumnos ejercía el papel de un *router* específico en la red, que debía entregar mensajes (en papel o acetato) a sus *routers* vecinos, y computar de manera asíncrona sus vectores de distancia y tablas de encaminamiento de acuerdo con la información recibida. Las dramatizaciones son un mecanismo muy potente para el aprendizaje de conceptos complejos, como puede ser la problemática de la cuenta a infinito. No obstante, su uso debe ser prudente, dado que son actividades que emplean mucho tiempo de la sesión.

A medida que avanzaba la asignatura y los alumnos se iban adaptando a la nueva metodología, se fueron incorporando actividades que demandaban mayor implicación por parte de los alumnos, pero que a su vez les daban un mayor control sobre lo que se hacía en las sesiones. Por ejemplo, a partir de la sesión 5, se pidió a los alumnos que nos enviaran posibles preguntas tipo test para hacer en las clases, similares a las que habíamos hecho en sesiones anteriores. Esto hacía que los alumnos tuvieran que enfrentarse al material de una forma activa, no sólo intentando comprender, sino también intentando extraer preguntas interesantes que pudieran suponer un desafío para sus compañeros. Este proceso, en general, implicaba varias iteraciones por correo electrónico entre profesor y alumno, y en ocasiones implicaba documentarse en fuentes externas, si la pregunta llegaba al límite del alcance de la asignatura (Figura 2).

Finalmente, siguiendo con el orden creciente de implicación en su aprendizaje, los alumnos tuvieron la oportunidad de preparar media sesión para sus compañeros en una de las sesiones finales de la asignatura. Los alumnos se dividieron en dos equipos, a cada uno de los cuales se les asignó la mitad de los contenidos de seguridad de la asignatura (divididos en criptografía y seguridad operacional). Durante la sesión 10, cada equipo comenzó a preparar, con ayuda del profesor, las actividades formativas adecuadas a los objetivos de aprendizaje de la siguiente sesión (relativas a los contenidos que se le habían asignado a su equipo). Los equipos terminaron la preparación de la sesión de manera autónoma como parte de su trabajo personal. Durante la sesión 11, cada equipo dio una hora de clase a sus compañeros sobre los contenidos correspondientes.

V. EVALUACIÓN DE LA EXPERIENCIA

A. Resultados académicos

Por supuesto, resulta arriesgado aventurar conclusiones significativas sobre la influencia del cambio metodológico en el rendimiento académico de los estudiantes, dada la escasa magnitud de la muestra. No obstante, en la Figura 3 (izquierda) se muestra la distribución de las calificaciones en una primera prueba parcial (PP1) del curso (en el momento de escritura de este artículo, aún no se dispone de las calificaciones definitivas de la segunda prueba parcial, ni se ha celebrado la prueba final). La figura muestra un diagrama de cajas de las calificaciones del total de alumnos presentados a la prueba (112 estudiantes), y en ella se han resaltado las calificaciones de los 10 alumnos que participaron en la experiencia. Hay que destacar que todos los alumnos de la asignatura se enfrentan a la misma prueba parcial, lo que permite una comparación directa (aunque no sea estadísticamente significativa). Los resultados sugieren un impacto positivo de la metodología sobre el rendimiento académico de los estudiantes, aunque esta hipótesis debería verificarse con una muestra más grande.

A modo de referencia, se muestran en la Figura 3 (derecha) los resultados académicos correspondientes a la misma prueba parcial en el curso anterior (2013/2014). Los

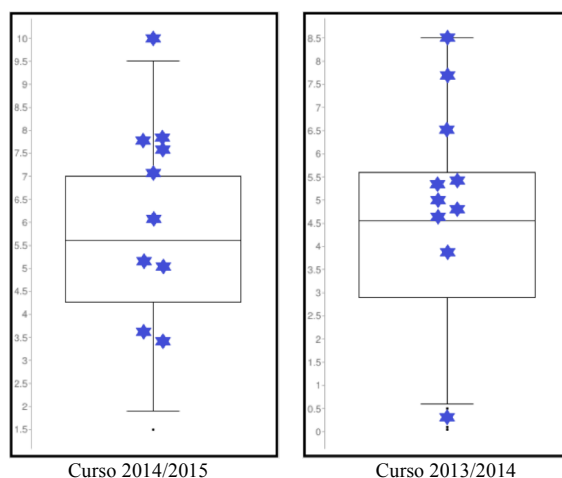


Fig. 3. Resultados académicos de la primera prueba parcial de AR2 en los cursos 2014/2015 y 2013/2014. Los valores resaltados son los que corresponden a los estudiantes del grupo que se imparte en inglés.

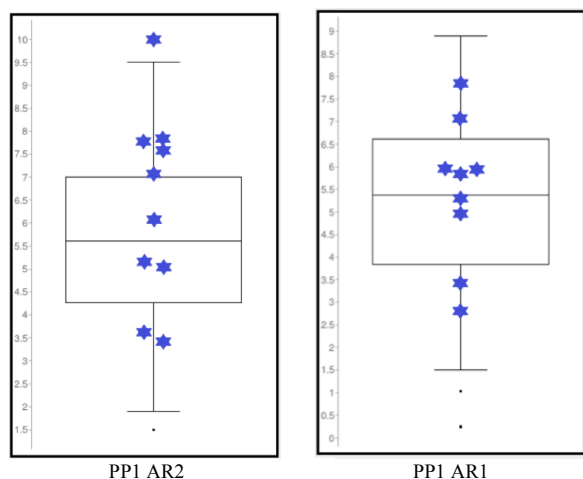


Fig. 4. Resultados académicos de la primera prueba parcial de las asignaturas AR1 y AR2 en los cursos 2014/2015.

ejes se han normalizado en tamaño para ofrecer una distribución visual independiente de los valores absolutos de las calificaciones (aunque se mantienen los valores numéricos originales en la leyenda). De este modo se permite observar la tendencia independientemente de la posible variación en la dificultad de la prueba en los diferentes años. Podemos observar que en la prueba parcial del año anterior también se observa una tendencia al alza en las calificaciones del grupo de inglés, por lo que esto podría deberse a un sesgo particular de dicho grupo (son alumnos que, voluntariamente, escogen un grupo bilingüe). De nuevo, se hace necesario repetir el experimento con una muestra más grande para sacar conclusiones significativas. En todo caso, para obtener una valoración independiente de ese posible sesgo, en la Figura 4 se muestran los resultados académicos de la PP1 del curso 2014/2015 de la asignatura AR1, donde participaron los mismos alumnos en el grupo de inglés, y donde se empleó una metodología tradicional de clases magistrales. Aunque de nuevo la muestra del grupo de inglés no es suficientemente grande para extraer conclusiones significativas, sí parece observarse que la distribución de las calificaciones de los alumnos del grupo de inglés de AR1 es bastante similar a la del grupo correspondiente de español (los alumnos se distribuyen de forma equitativa en los cuatro cuartiles). Sin embargo, en el caso de AR2, el grupo donde se realizó la experiencia tiene a la mitad de sus alumnos en el primer cuartil. Puesto que se trata de los mismos alumnos, este resultado sugiere que la metodología aplicada tiene un impacto positivo sobre su rendimiento académico, aunque de nuevo sería necesario repetir el experimento con una muestra más grande para poder verificar esta hipótesis.

Tabla I
ÍNDICES DE ASISTENCIA (CURSO 2014/2015)

Alumnos	Total alumnos (mat./pres.)	% asistencia primera mitad cuatrimestre (mat./pres.)	% asistencia segunda mitad cuatrimestre (mat./pres.)
Español	144 / 112	40 / 51	33 / 42
Inglés	12 / 10	83 / 100	83 / 100

B. Asistencia a clase de los alumnos

Otra medida relevante del éxito de la experiencia puede ser el nivel de asistencia de los alumnos, bajo la premisa de que los alumnos acuden a clase si tienen la percepción de que eso les resulta útil. Por otro lado, uno de los factores que impulsaron la realización de esta experiencia fue la percepción de que las clases de grupo grande no atraían al alumnado, percepción derivada (entre otros factores) de los bajos índices de asistencia.

En la Tabla I se muestran los índices de asistencia en porcentaje tanto de los grupos de español como del de inglés en el curso 2014/2015, tanto con respecto al número total de matriculados (*mat.*) como frente al número de alumnos que se presentaron a la primera prueba parcial (*pres.*). Esta última distinción es relevante, ya que los alumnos tienen la posibilidad de renunciar a la evaluación continua si por motivos laborales no pueden seguir el proceso de evaluación continua (adhiriéndose a un sistema de evaluación final). Podemos ver una relación similar de presentados frente a matriculados en ambas modalidades de la asignatura. Sin embargo, los porcentajes de asistencia varían de forma apreciable. En particular, podemos ver que el total de alumnos que siguen la evaluación continua en el grupo de inglés asisten a clase regularmente, mientras que la proporción en los grupos de español es de en torno al 50%, y decrece a medida que el curso avanza. Esto podría ser un indicador de que la metodología aplicada en la experiencia hace las clases más provechosas (al menos desde un punto de vista subjetivo) para los alumnos. No obstante, carecemos de datos de asistencia del curso anterior, por lo que no podemos descartar la posibilidad de que se trate de un sesgo por las particularidades del grupo de inglés.

C. Encuestas y valoraciones subjetivas

Durante el transcurso del curso, y siguiendo con la premisa de la realimentación temprana de la que hablábamos más arriba, se realizaron una serie de encuestas anónimas y voluntarias a través de la plataforma Blackboard y de la herramienta Socrative. En estas encuestas se recababa la opinión del alumno sobre distintos aspectos de la experiencia, en general empleando escalas Likert [10], en las que el alumno tenía que valorar de 1 a 5 lo de acuerdo o en desacuerdo que estaba con cada una de las afirmaciones que se le presentaban.

En primer lugar, estábamos interesados en saber hasta qué punto los diferentes elementos de la metodología estaban funcionando. Era crítico para el éxito de la experiencia que los alumnos efectivamente trabajasen el material de la asignatura antes de venir a clase, y también que la clase

Tabla II
ENCUESTAS RESPECTO AL FUNCIONAMIENTO DE LA METODOLOGÍA

Pregunta	Valoración media (1-5)
I received adequate instructions on how to prepare before the class	4.14
I prepared before the class according to the professor's instructions	3.86
Preparing for the class beforehand helped me to make the most of the session	4.16
The session facilitated my understanding of the concepts I previously prepared about	4.42
After my personal work and the LG session, I think I was adequately prepared for the SG session that followed.	4.00

aportase algo adicional sobre el aprendizaje autónomo del alumno. Por otro lado, era necesario garantizar que la experiencia no impactaba negativamente a la parte de grupo pequeño de la asignatura. Las preguntas relacionadas con estos aspectos, junto con la opinión media del alumno sobre las mismas, se muestran en la Tabla II. Podemos ver que las valoraciones son favorables, y que con lo que los alumnos se muestran más críticos es con su propio compromiso con la preparación previa.

Especialmente durante las primeras sesiones, se incluyeron preguntas para valorar el impacto emocional de la nueva metodología en los alumnos, dado que ésta implicaba exigir a los estudiantes una mayor participación, e incluso la participación en actividades inusuales. Dichas preguntas se formularon de forma similar a los *emotion assessments* que

Tabla III
VALORACIÓN EMOCIONAL DE LA EXPERIENCIA

Pregunta	Valoración media (1-5)
What is the most _____ you felt during class?	
amused/fun	4.27
dismissed/disrespected/ignored	1.00
embarrassed/awkward/ridiculous	1.00
engaged/interested/curious	3.89
clueless/lost	2.00
self-assured/capable	3.50

Tabla IV
VALORACIÓN DE LA ADECUACIÓN DE LA METODOLOGÍA

Pregunta	Valoración media (1-5)
The activities we did in class were adequate	4.45
In general, attending to class was worth the time	4.19
In general, I think I am learning more than I would with traditional masterclasses	4.43

Tabla V
VALORACIÓN DE ACTIVIDADES ESPECÍFICAS

Pregunta	Valoración media (1-5)
Preparing my questions for the class helped me understand better this (or other) week contents	4.00
My submitted question was presented/discussed adequately during the class	4.20
Having to teach specific topics to my classmates has helped me to better understand these topics	4.00
Having my classmates teaching me course contents has helped me to better understand these concepts	3.33

Tabla VI
VALORACIÓN FINAL DE LA EXPERIENCIA

Pregunta	Valoración media (1-5)
In general, attending to the sessions has been a rewarding experience.	4.75
I have probably learned more than I would have with traditional masterclasses	5.00
I think the methodology of the sessions in this course may have had a negative impact on my academic performance (i.e. my grades)	1.00
If I had a friend taking the AR2 course next year, I'd recommend him/her to take the English group.	5.00
I'd be glad to find a similar methodology in future courses of this degree	4.75

se hacen en psicología [11], donde se valoran hasta qué punto se han experimentado las emociones aludidas en una escala de 1 (en absoluto) a 5 (extremadamente). La Tabla III muestra las valoraciones de los alumnos a este respecto. Podemos ver cómo las valoraciones son altas para las emociones positivas y bajas para las negativas. Especialmente relevante es el hecho de que ningún alumno se sintió en ningún momento ridículo (algo que podría haber pasado en las dramatizaciones o en las presentaciones) o no respetado adecuadamente (algo que podría haber pasado con las aportaciones de los alumnos durante la clase).

Una vez el curso iba avanzando y los alumnos se iban adaptando a la metodología, dimos la oportunidad a los alumnos de opinar acerca de la adecuación de la misma para el aprendizaje. Los resultados de estas valoraciones pueden verse en la tabla IV, donde se observa que los alumnos valoran positivamente la aportación de las clases presenciales a su aprendizaje.

También se introdujeron en diferentes momentos del curso preguntas para que valorasen cuáles de las actividades realizadas contribuían mejor a facilitar su aprendizaje. La Tabla V muestra, a modo de ejemplo, las relativas a las preguntas enviadas por los alumnos y a las sesiones en las que fueron los estudiantes los que dieron clase a sus compañeros. De nuevo las valoraciones son positivas, aunque destaca el hecho de que de forma generalizada los alumnos no se mostraron especialmente satisfechos con las clases recibidas de sus compañeros, pero sí con impartirlas.

Finalmente, al término de la asignatura, se realizó una encuesta de valoración global a modo de retrospectiva. Como puede verse en los resultados de dicha encuesta (Tabla VI), las valoraciones globales son muy positivas. Cabe destacar el hecho de que la opinión generalizada de los alumnos es que la metodología les ha hecho aprender más que en clases magistrales, y que no ha tenido un impacto negativo en sus resultados académicos.

VI. CONCLUSIONES

En este artículo se ha presentado una experiencia de *flipped classroom* en el contexto de una asignatura básica de redes. La experiencia se ha diseñado alrededor de los principios generales de reducir las exposiciones tipo clase magistral a breves introducciones a los temas que los alumnos debían preparar por su cuenta, para luego trabajarlos en clase mediante actividades de aprendizaje activo, tales como preguntas para la discusión (apoyadas por herramientas como *Socratic*) y dramatizaciones. Asimismo, de manera incremental los alumnos han tenido la oportunidad de ir adquiriendo protagonismo en su proceso de aprendizaje, siendo ellos los que ideaban parte de las preguntas que se planteaban en clase, e incluso llegando a ser ellos (por equipos) los que actuaran como docentes en una de las últimas sesiones de la asignatura.

Los resultados de la experiencia, tanto desde el punto de vista de las calificaciones obtenidas por los alumnos como de la valoración de la experiencia por parte de los mismos mediante encuestas han sido muy satisfactorios. No obstante, existen numerosos aspectos de mejora para el futuro. En primer lugar, como hemos visto en el análisis de resultados, el reducido tamaño de la muestra impide extraer conclusiones estadísticamente significativas, por lo que uno de nuestros intereses más inmediatos es repetir la experiencia

con un grupo más grande. Esto plantea desafíos adicionales, tanto para el trabajo en clase como para la evaluación. Uno de los problemas de hacer dramatizaciones y otras actividades de carácter grupal con un número de alumnos elevados, es la dificultad para el profesor de atender debidamente a todos los grupos. Una posible solución sería escoger en cada sesión un conjunto reducido de alumnos que trabajen durante la semana con el profesor, y luego ayuden a sus compañeros en las actividades de la sesión propiamente dicha. A la hora de valorar la participación en clase, puede emplearse la capacidad de *Socratic* de realizar tests no anónimos, aunque quizás eso introduzca un factor de presión sobre los alumnos que anule la atmósfera de experimentación durante la clase. Otra posibilidad sería buscar un esquema de *peer review* entre los estudiantes.

También hay aspectos de la metodología de *flipped classroom* en los que podríamos profundizar más. Por ejemplo, en esta experiencia no se ha hecho mucho hincapié en actividades *post*-sesión, y las actividades *pre*-sesión se han limitado a lectura, estudio y resolución de ejercicios. Podrían plantearse actividades opcionales de ampliación sobre determinados contenidos, y diversificar el material que se pone a disposición de los alumnos (incluyendo, por ejemplo, videos). Esto además contribuiría a cubrir otra de las “buenas prácticas” de las que se hablaba en [1]: respetar la diversidad de talentos y proporcionar diferentes vías de aprendizaje.

Finalmente, es necesario calibrar el esfuerzo del docente en la preparación de este tipo de asignaturas. En este caso concreto, se trataba de una asignatura con cierto recorrido que contaba con un libro de texto que se seguía fielmente y material de muy alta calidad, lo que ha reducido en gran medida la labor de preparación de material de referencia para los alumnos. La preparación de una asignatura “desde cero” sin un texto de referencia sería previsiblemente más costosa.

AGRADECIMIENTOS

Esta trabajo se ha realizado en el contexto de un proyecto de innovación docente de la “Convocatoria de proyectos para el fomento de la innovación en el proceso de Enseñanza-Aprendizaje”, perteneciente al Programa Propio de la UAH.

REFERENCIAS

- [1] A.W. Chickering, A. W. y Z.F. Gamson, “Seven principles for good practice in undergraduate education”. *Biochemical Education*, vol. 17, no. 3, pp. 140–141, 1989.
- [2] Lage, Maureen J., Glenn J. Platt, and Michael Treglia. "Inverting the classroom: A gateway to creating an inclusive learning environment." *The Journal of Economic Education* 31, no. 1 (2000): 30-43.
- [3] Mazur, Eric. "Farewell, lecture?" *Science* 323, no. 5910 (2009): 50-51.
- [4] Slavin, Robert E. "Evidence-based education policies: Transforming educational practice and research." *Educational researcher* 31, no. 7 (2002): 15-21.
- [5] Tucker, Bill. "The flipped classroom." *Education Next* 12, no. 1 (2012): 82-83.
- [6] Bishop, Jacob Lowell, and Matthew A. Verleger. "The flipped classroom: A survey of the research." In *ASEE National Conference Proceedings, Atlanta, GA*. 2013.
- [7] James F. Kurose y Keith W. Ross. “Computer networking: a top-down approach”, 6ª Ed, Addison-Wesley Publishing Company, 2012.
- [8] Benjamin S. Bloom. “Taxonomy of Educational Objectives Book 1: Cognitive Domain”, Addison-Wesley Publishing Company, 1984.

- [9] Hung, Woei, David H. Jonassen, and Rude Liu. "Problem-based learning." *Handbook of research on educational communications and technology* 3 (2008): 485-506.
- [10] Likert, R., “A technique for the measurement of attitudes. *Archives of Psychology*”, vol. 22, n. 140, pp. 1-54, 1934.
- [11] Fredrickson, Barbara L. "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions." *American psychologist* 56, no. 3 (2001): 218.

LIDER: experiencia para mejorar la eficiencia del aprendizaje colaborativo en grupos de tamaño medio

Guillermo Azuara¹, M^a Carmen Aguilar², Julián Fernández-Navajas¹, José Luis Salazar¹,
Inmaculada Plaza¹ y José Luis Tornos¹.

¹Departamento de Ingeniería Electrónica y Comunicaciones,

²Departamento de Derecho de la Empresa

Universidad de Zaragoza

Pedro Cerbuna, 12. 50009 Zaragoza - España.

{gazuara, caguilar, navajas, jsalazar, inmap, jltornos}@unizar.es

Resumen- Las actuales tendencias educativas, con el alumno como centro del proceso aprendizaje, hacen que sea habitual el uso de metodologías activas como recurso docente. En este contexto, se detectó un problema de motivación por parte de algunos alumnos, en el seguimiento de actividades de aprendizaje colaborativo, en grupos de tamaño medio (de 3 a 8 miembros). LIDER (LIDÉrazgo Rotatorio) surgió como una experiencia para intentar aumentar la motivación de los alumnos, basado en que el rol de líder del grupo va rotando en cada práctica, reforzando la interdependencia de los miembros del grupo, y la responsabilidad de todos y cada uno de los alumnos.

Palabras Clave- aprendizaje colaborativo, aprendizaje cooperativo, metodologías activas, problemas de motivación, grupos de tamaño medio.

I. INTRODUCCIÓN

Con la implantación del Espacio Europeo de Educación Superior, el estudiante se ha convertido en el centro del proceso educativo. La labor del profesor se va alejando más de la mera clase magistral, y tiende a centrarse en una adecuada planificación de la asignatura, utilizando las metodologías que crea más acertada para cada materia, o incluso partes de materias. Los profesores no sólo deben conocer bien la materia que han de enseñar, sino también deben conocer cómo aprenden los estudiantes cuyo aprendizaje han de guiar y, además, conocer los recursos de enseñanza para poder llevar a cabo su trabajo, entre los que destacan las metodologías activas [1]. Los profesores que utilizan enfoques de aprendizaje colaborativo se convierten en diseñadores de experiencias intelectuales, en un elemento más del proceso de aprendizaje en el que se establece una relación de interdependencia entre profesor y alumnos [2].

De entre las metodologías activas [1], una de las más completas y que tiene más situaciones en que puede ser aplicada es la del aprendizaje cooperativo. Como se indica en [3]: “Aprender es algo que los alumnos hacen, y no algo que se les hace a ellos. El aprendizaje no es un encuentro deportivo al que uno puede asistir como espectador. Requiere la participación directa y activa de los estudiantes. Al igual que los alpinistas, los alumnos escalan más fácilmente las cimas del aprendizaje cuando lo hacen formando parte de un equipo cooperativo.

La cooperación consiste en trabajar juntos para alcanzar objetivos comunes. En una situación cooperativa, los individuos procuran obtener resultados que sean beneficiosos para ellos mismos y para todos los demás miembros del grupo. El aprendizaje cooperativo es el empleo didáctico de grupos reducidos en los que los alumnos trabajan juntos para maximizar su propio aprendizaje y el de los demás (...) el docente puede organizar cooperativamente cualquier tarea didáctica, de cualquier materia y dentro de cualquier programa de estudios”.

La experiencia nos dice que con el uso de esta metodología, es habitual en los grupos de trabajo de tamaño mediano (de 3 a 8 miembros), que algunos estudiantes, bien por su naturaleza, bien por sus conocimientos o normalmente por la convergencia de ambas cosas, ejerzan como líderes del grupo. Esto hace que determinados estudiantes queden, a veces, en un segundo plano o incluso relegados de tareas muy importantes para la adquisición de competencias, lo que hace que en ocasiones, a pesar de tener una buena nota en la parte de trabajo con el grupo, a nivel individual ni siquiera alcancen los mínimos para superar la asignatura. Esto significa que hemos fracasado en nuestra labor de facilitadores de adquisición de conocimiento, y en la mayoría de las ocasiones esto no se detecta hasta la realización de pruebas individuales, donde se aprecia una gran diferencia entre la nota individual del alumno y la nota grupal, con el agravante de que en el caso de prácticas complejas esta detección suele ser demasiado tardía para aplicar medidas correctoras.

Con esta experiencia, se espera mejorar la eficiencia del aprendizaje (mejor aprovechamiento del tiempo), asociada a una mayor motivación, derivada de la mayor implicación emocional de los alumnos con el grupo tanto por ayudar al grupo como, desde un punto de vista más egoísta, porque cuando ellos sean líderes necesitarán del apoyo del grupo. Con ello conseguimos aumentar la capacidad de liderazgo, comunicación, trabajo en grupo y el mejor aprovechamiento de las horas de prácticas, mediante el aumento de la implicación del grupo.

El intercambio de opiniones de profesores sobre los resultados del proyecto en grupos de alumnos muy distintos,

puede generar una imagen bastante fidedigna de la utilidad real del método propuesto.

La propuesta de la experiencia LIDER (LIDERazgo Rotatorio) se plantea de forma que en cada sesión de prácticas uno de los alumnos (designado por el profesor) deberá adoptar el rol de líder. Éste se encargará de repartir las tareas, planificar los tiempos y realizar una preparación previa especialmente exhaustiva de esa práctica. Aunque él se prepare muy bien esa práctica, sus compañeros también deberán preparar muy bien la parte que les asigne el líder, y tendrán que ayudar con sus conocimientos específicos al resto de compañeros, a alcanzar las competencias y habilidades marcadas como objetivo en cada práctica.

Obviamente, esto conlleva el uso de unas técnicas docentes adecuadas (en nuestro caso la coevaluación y la autoevaluación), que se describirán posteriormente.

En el resto del artículo se presentará cómo se ha diseñado la experiencia, los métodos, actividades y tecnologías empleadas, el desarrollo de la experiencia, resultados y finalmente las conclusiones y trabajo futuro.

II. DISEÑO DE SESIONES LIDER

En las metodologías activas, una de las partes más importantes del proceso de aprendizaje es el adecuado diseño del proceso docente por parte del profesor. De la adecuada selección de objetivos, tareas, temporización y motivación de los alumnos, va a depender en gran manera el éxito del aprendizaje, materializado como adquisición de competencias y comprobable mediante una adecuada evaluación.

Hay abundante bibliografía centrada en la utilización en el aula de metodologías activas, y también dentro de éstas en el aprendizaje colaborativo. Como se indica en [4], *“aprendizaje colaborativo describe una situación en la cual se espera que ocurran formas particulares de interacción, que conllevarán a mecanismos de aprendizaje, pero no hay una garantía total que éstas condiciones ocurran efectivamente. (...) La colaboración solamente podrá ser efectiva si hay una interdependencia genuina entre los estudiantes que están colaborando. Esa dependencia genuina la describe como:*

1.) *La necesidad de compartir información, llevando a entender conceptos y obtener conclusiones.*

2.) *La necesidad de dividir el trabajo en roles complementarios*

3.) *La necesidad de compartir el conocimiento en términos explícitos [5]”.*

Como paso previo, deberemos tener claro cómo preparar una sesión de trabajo con esta metodología. La primera parte de una sesión LIDER, consiste en la adecuada preparación de las sesiones, siempre teniendo en cuenta las pautas metodológicas asociadas a la herramienta docente elegida. Por ejemplo en la guía de Van Til & Van der Heidjen [6], señala las siguientes actividades:

- Acciones Previas.
- Definir los objetivos.
- Definir el tamaño del grupo.
- Definir la composición del grupo.
- Definir la distribución en el laboratorio.
- Definir los materiales de trabajo.
- Dividir el tema y objetivos en subtareas.

- Lluvia de ideas respecto al tema.

En [7] se presenta una lista de tareas muy similares a la anterior.

Una vez diseñada la actividad, y formados los grupos, el profesor responsable de cada asignatura seleccionará un líder por grupo para cada práctica, lo comunicará con la suficiente antelación y proporcionará a todo el grupo la documentación de dicha práctica (se incluirá un guión de la práctica, un entregable y los formularios de valoración que luego deberán completar). Por tanto LIDER requiere un trabajo extra, consistente en designar un líder para cada sesión (asegurando que al menos cada uno de los alumnos sea líder de su grupo en una sesión de prácticas), preparando las rúbricas y haciendo un seguimiento de cómo se ejerce el liderazgo.

Las rúbricas se han diseñado todas con tres niveles (malo – regular – bueno), para cada uno de los aspectos a valorar o indicadores.

Además, deberá transmitirse de forma clara a los alumnos las tareas a realizar como líderes del grupo: coordinar a sus compañeros para la consecución de los objetivos en el tiempo marcado, reparto de tareas, puesta en común y resolución en el laboratorio del reto planteado en esa sesión de prácticas.

También se comunicará tanto al líder como a sus compañeros la forma de evaluación. Se les proporcionará tanto el formulario para evaluar como las rúbricas necesarias para la coevaluación y para la autoevaluación. Entre otras ventajas, con el uso de rúbricas conseguiremos homogeneizar los criterios de evaluación tanto dentro del grupo, como entre grupos. La experiencia demuestra que el sólo hecho de anunciar esta valoración grupal e individual, aumenta el rendimiento del grupo.

Es necesario además, preparar una rúbrica específica para hacer un seguimiento del desempeño del líder en su rol. Por supuesto, también se dará a conocer a los alumnos para que sepan que van a ser evaluados en su tarea de líder y cómo va a ser esa evaluación.

III. MÉTODOS, ACTIVIDADES Y TECNOLOGÍAS UTILIZADAS.

A. Métodos.

Para la realización de las prácticas, además de una pequeña introducción por el profesor, y la utilización de guiones (autoaprendizaje) y formularios para completar (entregables), se han utilizado los siguientes recursos docentes:

- LIDER: es el método que se pretende desarrollar y probar.
- AUTOEVALUACIÓN y COEVALUACIÓN [4]: se prepararán las rúbricas y las herramientas que permitan a los alumnos desarrollarlas eficazmente.
- SEGUIMIENTO [5]: el profesor también realizará un seguimiento de la forma en que cada líder desarrolla su tarea.

B. Actividades.

- Preparación de los guiones de prácticas y rúbricas para su adaptación a LIDER.

- Reuniones con los profesores implicados en la experiencia, para intercambiar información sobre el desarrollo de la experiencia en diferentes entornos.
- Encuesta a los alumnos sobre la metodología.
- Encuesta a los profesores sobre la metodología.

C. Tecnologías empleadas.

Como soporte tecnológico de la experiencia, es necesario un sistema fiable y rápido de recogida de datos, especialmente de las autoevaluaciones y coevaluaciones de los alumnos. Como la evaluación que realizan los alumnos será tomada en cuenta para la nota final, será necesaria una autenticación de los usuarios. Para las encuestas de valoración de la actividad, se trabajará con encuestas anónimas de los alumnos. Para realizar esta toma de datos se barajaron dos alternativas, el uso de formularios de Google o el uso de una plataforma propia.

El uso de los formularios de Google tiene como ventajas que su manejo es muy sencillo e intuitivo, y muchos usuarios están acostumbrados a interactuar con productos de Google.

Las ventajas que pueden suponer el uso de una plataforma propia con autenticación de firmas en anillo (web o en versión de aplicación para móvil), en comparación con los formularios Google son muy diversas. Nuestra plataforma implementa un sistema de seguridad criptográfica basado en firmas en anillo. Esto proporciona al sistema todas las características de seguridad propias de una e-votación: confidencialidad, democracia, anonimato (si fuese requerido), etc. que ayudan a generar un clima de seguridad y confianza al alumno que fomente su participación. Por otra parte, se puede hacer un seguimiento a través de diferentes consultas de la evolución de pensamiento de una persona, manteniendo el anonimato, pudiendo detectar síntomas de liderazgo si las ubicamos en la línea temporal junto con actividades en la vida real que hayan podido influir en el cambio de opinión o pensamiento por actuación de una persona (posible líder). Por otra parte, el uso de una tecnología tan de moda como son las aplicaciones móviles la hacen más atractiva a los alumnos [8].

IV. DESARROLLO DE LA EXPERIENCIA

Inicialmente la experiencia estaba pensada para las siguientes asignaturas:

- Grado de Ingeniería Informática: Diseño y Administración de Redes (4º curso) (Obligatoria / Optativa), Prevención de Riesgos Laborales aplicada a la ingeniería (4º curso) (Optativa).
- Grado de Ingeniería Electrónica y Automática: Prevención de Riesgos Laborales aplicada a la ingeniería (4º) (Optativa).
- Grado en administración y dirección de empresas: Introducción al Derecho (1º) (Formación Básica).

Antes de iniciar las prácticas el profesor o grupo de profesores prepararán el material que se ha comentado en el punto II. También se deberá fijar el peso que va a tener en la nota las calificaciones que van a emitir los alumnos, para hacérselo saber a estos antes del inicio de la actividad.

Durante la práctica además del seguimiento usual de la práctica, el profesor realizará un seguimiento de cómo el

líder de cada grupo desempeña su rol. Los alumnos por su parte irán realizando la práctica y completando un entregable que enviarán al profesor al acabar la sesión (o en un plazo determinado si se considera más conveniente).

Al finalizar la práctica se pedirá a los alumnos que introduzcan sus valoraciones, sobre sí mismos, sus compañeros y qué tal ha desempeñado el líder su función (autoevaluación y coevaluación), mediante un formulario electrónico, y recordando que es muy importante que les hayamos facilitado con antelación una rúbrica para la evaluación.

El profesor calificará el entregable y estudiará las diferentes puntuaciones indicadas por los alumnos. Haciéndoles llegar la corrección con las observaciones pertinentes.

Además, tras la realización de la experiencia, los profesores implicados en el proyecto se reunirán para realizar un seguimiento del proyecto y finalmente para estudiar los datos obtenidos y analizarlos para extraer las conclusiones oportunas.

El calendario planteado para el desarrollo de la experiencia fue el siguiente:

- Enero: diseño de cuestionarios y soporte para la recogida de datos.
- Febrero: utilización del método en algunos de los grupos (hasta Mayo).
- Abril: reunión de seguimiento.
- Mayo: tratamiento de datos y conclusiones.

V. RESULTADOS.

Al principio de la experiencia se detectó, que con los alumnos de primer curso era muy difícil el desarrollo de la experiencia, por lo que se descartó trabajarla en asignaturas de primeros cursos.

Por cuestiones personales de uno de los profesores participantes, estuvo de baja, no fue posible realizar la experiencia en dos de las asignaturas, por lo que nos vimos obligados a posponer al próximo curso la realización de una experiencia más amplia, tal y como había sido diseñada.

Los objetivos esperados eran conseguir una mayor motivación de los alumnos, sobre todos de los que tienen más dificultad en seguir la asignatura, que redunde en una mejora del proceso aprendizaje.

La principal innovación en el sistema propuesto, es que combina metodologías bien conocidas (aprendizaje colaborativo, coevaluación y autoevaluación) con la designación rotatoria del líder del grupo por parte del profesor.

Para comprobar las mejoras, se había planteado realizar una o dos sesiones sin aplicar LÍDER, pero intentando cuantificar el nivel de motivación (mediante encuesta) y aprendizaje de los alumnos, para poder contrastarlo con las prácticas en las que se aplique el método propuesto.

Otra alternativa, si el número de grupos de prácticas lo permitía, era aplicarse en todos los grupos menos en uno, que serviría como grupo de control para poderlos comparar a ver si se aprecia mejora significativa.

Para completar la evaluación se analizarán tanto los resultados académicos (comparándolos con otras sesiones de prácticas y otros grupos), como las encuestas de satisfacción

institucionales que completarán los alumnos al finalizar el periodo docente.

Debido al bajo número de alumnos (se trata de asignaturas de 4º curso, de titulaciones de reciente implantación, es el segundo curso que se imparten), no ha sido factible llevar a cabo las comparaciones previstas, por lo que sólo se ha contado con las respuestas de una reducida muestra de 8 alumnos.

La encuesta de valoración de la actividad constaba de preguntas tipo test. Se muestran a continuación:

Valora de 1 a 5 (1 es la peor valoración y 5 la mejor).

1. Planificación de la actividad.
2. Exposición clara de la mecánica de la actividad.
3. Documentación facilitada.
4. Exposición de los criterios de evaluación.
5. Logro de los objetivos planteados.
6. Valoración general de la actividad.

Valora de 1 a 5 (1 totalmente en desacuerdo y 5 totalmente de acuerdo).

7. La actividad me ha hecho preparar mejor la práctica (trabajo previo).
8. La actividad ha conseguido que el trabajo se reparta más equitativamente entre los miembros del grupo.
9. Cada miembro del grupo había preparado su parte de la tarea adecuadamente.
10. En general, el líder en cada práctica ha ejercido como tal coordinando la realización de la tarea.
11. Creo que ha ayudado a que algunos compañeros se enganchen más con la asignatura.
12. Creo que con la actividad ha mejorado el ambiente en el grupo.

Respuestas abiertas:

13. ¿Qué valoras como más positivo de la actividad?
14. ¿Qué valoras como más negativo de la actividad?
15. Si deseas hacer algún comentario o sugerencia, hazlo a continuación.

Los resultados de la encuesta se muestran en las figuras 1 a 4.

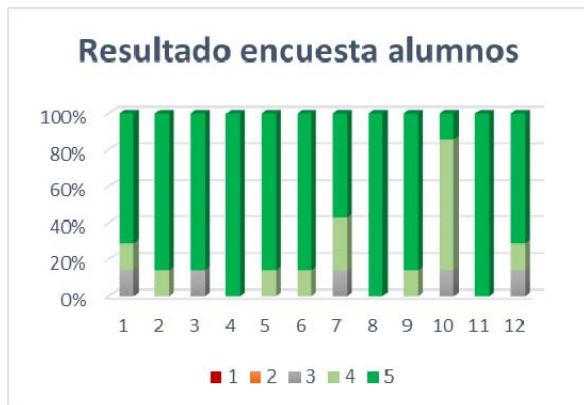


Fig. 1. Resultado de encuestas a alumnos.

Como se puede apreciar en la Fig.1 en general la valoración de la actividad por parte de los alumnos ha sido

muy positiva, no calificando ninguna de las preguntas por debajo de 3, y siendo la mayoría de las valoraciones 4 o 5.

Respecto a la valoración de la actividad (Fig. 2), los alumnos la han valorado muy bien, correspondiendo una amplia mayoría (86 %) a la máxima valoración (5 sobre 5) y el resto (14 %) a la siguiente más alta (4 sobre 5).

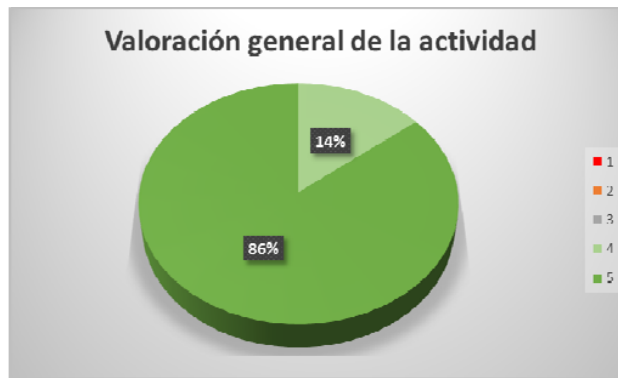


Fig. 2. Valoración general de la actividad.

Uno de los objetivos, que era conseguir aumentar la implicación de todos los alumnos del grupo, a la vista de la Fig. 3 se ha conseguido (en opinión de los alumnos). Una pregunta muy similar, la 11 (Creo que ha ayudado a que algunos compañeros se enganchen más con la asignatura), ha presentado los mismos resultados, por lo que parece que la experiencia sí ha servido para recuperar a los alumnos más “descolgados”.

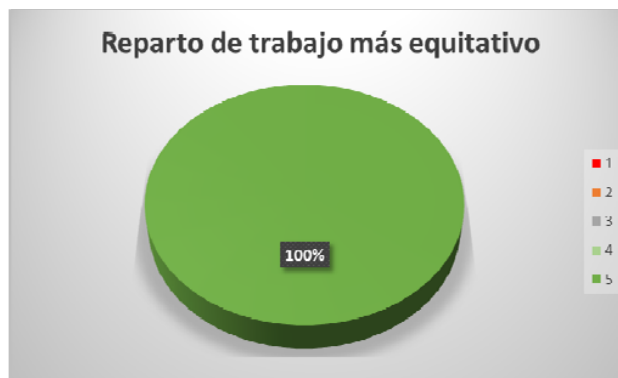


Fig. 3. Reparto del trabajo equitativo.

La pregunta que ha obtenido menos valoración ha sido la que preguntaba sobre si se creía que este tipo de actividad mejoraba la preparación previa de la práctica. Aunque hay una mayoría de respuestas positivas (86 %), ver Fig. 4, nos parece preocupante, ya que uno de los objetivos era también que se mejorara el trabajo previo. La respuesta podría obedecer a que los estudiantes que ya hacían una buena preparación previa de las prácticas, este método no les hace mejorar en este aspecto (porque ya no era susceptible de mejora). El próximo curso se añadirá una pregunta de autovaloración sobre el trabajo previo de los alumnos.



Fig. 4. Preparación previa de la práctica.

Respecto a las preguntas abiertas, destacar que algunos de los alumnos han demandado una formación sobre cómo desarrollar adecuadamente el rol de líder. Aunque los autores creíamos que la rúbrica por sí misma era suficiente orientación, estas respuestas nos hacen plantearnos la necesidad de aportar algún tipo de formación transversal en este sentido (fuera de las asignaturas).

También se ha señalado que en el caso de prácticas de diferente nivel de dificultad, no parecía muy equitativa la valoración de los líderes. Se trabajará este aspecto para intentar mantener la máxima equidad entre los alumnos.

Algunos hablan que por respeto a sus compañeros, y por recibir el mismo trato cuando ellos sean líderes, creen que todos los miembros del grupo han trabajado más.

También se han encontrado varias respuestas en el sentido de que los estudiantes agradecen el uso de metodologías activas (los estudiantes lo denominaban en sus respuestas "este tipo de prácticas").

Finalmente, otro número importante de alumnos pedía más tiempo para las prácticas. Esta cuestión depende de la planificación de las asignaturas, y aunque se revisará, se ha detectado en algunas ocasiones que el trabajo previo de los alumnos en la preparación de las prácticas no es el más adecuado. En esta línea ya se ha trabajado en alguna ocasión, haciendo que antes de comenzar la práctica el alumno deba presentar un entregable con el trabajo previo encomendado.

VI. CONCLUSIONES Y TRABAJO FUTURO.

A pesar de todas las dificultades y limitaciones que hemos tenido para llevar a cabo esta primera experiencia, y siendo por tanto su validez muy limitada por el reducido tamaño y representatividad de la muestra, sí que se han extraído algunas conclusiones:

- Rotar el papel de líder, en opinión de los alumnos, hace que por respeto a sus compañeros se involucren más en la asignatura, por lo que podríamos inferir que se mejora el nivel de motivación.
- La aplicación de este método, y probablemente la mayoría de metodologías de trabajo colaborativo o cooperativo, parece más apropiada para los últimos cursos de la titulación.

- Dado que supone una carga de trabajo importante para el profesor, hay que valorar si es posible su aplicación en una asignatura.
- Todo el trabajo de planificación, documentación entregada por el profesor y uso de metodologías activas, es percibida de manera muy positiva por los alumnos.
- Se ha conseguido el objetivo principal que era intentar aumentar la motivación de los alumnos en la realización de prácticas en grupos de tamaño medio (donde la responsabilidad queda más diluida en el grupo).
- Hay que tener en cuenta que haya prácticas suficientes para que todos los alumnos puedan ejercer el rol de líder, por lo que hay que planificar las sesiones de prácticas teniendo esto en cuenta.

Con el uso de la plataforma propia para la recogida de datos mediante formularios electrónicos, hay una cuestión no trivial que es hacer llegar a los alumnos sus firmas electrónicas y sobre todo concienciarlos para que las custodien adecuadamente.

Aunque finalmente, la experiencia no ha podido ser tan ambiciosa en cuanto a número de asignaturas y número de alumnos como se había planteado inicialmente, todo el trabajo realizado, así como la reflexión sobre la experiencia se consideran de forma muy positiva, y permiten que se esté en muy buena disposición de cara a repetir la experiencia en el próximo curso.

También, de cara al próximo curso, se ha planteado la posibilidad de contar en el equipo docente con un experto en psicometría, con el objetivo de orientar las encuestas más específicamente, para poder obtener el máximo número de indicadores de cara a su posterior análisis y facilitar la comparación entre entornos heterogéneos.

AGRADECIMENTOS

Este proyecto ha sido financiado por el Programa de Incentivación de la Innovación Docente en la UZ (PIIDUZ) dentro de la Convocatoria de Innovación Docente 2014-2015 de la Universidad de Zaragoza, con código PIIDUZ_14_395.

REFERENCIAS

- [1] Amparo Fernández March, "Metodologías activas para la formación de competencias", *Educatio siglo XXI*, 24, pp. 35-56, 2006.
- [2] Sempere Ortells, J.M.; García Irlas, M.; Marco de la Calle, F.M.; De la Sen Fernández, M.L., "Aprendizaje colaborativo: un reto para el profesor en el nuevo contexto educativo". *El trabajo colaborativo como indicador de calidad del Espacio Europeo de Educación Superior*, vol 1, pp. 89-104, 2011.
- [3] Johnson, David W.; Johnson, Roger T.; Holubec, Edythe Johnson. *El aprendizaje cooperativo en el aula*. Barcelona: Paidós, 1999.
- [4] Collazos, César, Luis Guerrero, and Adriana Vergara, "Aprendizaje Colaborativo: Un cambio en el rol del profesor." *Proceedings of the 3rd Workshop on Education on Computing*, Punta Arenas, Chile. 2001.
- [5] Salomon Salomon, G., "What does the design of effective CSCL require and how do we study its effects?" *SIGCUE Outlook*, Special Issue on CSCL, 21(3), 62-68, 1992.
- [6] Til, Van C.T & Heijden, van der F. *Study skills in problem based learning*. Universitaire Pers, Maastricht, 1996.

- [7] Jonson D., Jonhson R, Jonhson E. The nuts and bolts os cooperative learning. Interaction Book Company, Minnesota, 1994
- [8] J.L. Salazar, J.J. Piles, J. Ruiz, J.M. Moreno, "Security approaches in e-cognocracy". Computer Standards and Interfaces, vol. 32, n. 5-6, pp. 256-265, 2010.

Gamificación en el aula universitaria: un caso práctico en una asignatura de Redes

Diego Fernández, Fidel Casheda, Francisco Nóvoa de Manuel,
Xacobe Macía da Silva, Victor Carneiro

Departamento de Tecnologías de la Información y las Comunicaciones,
Universidad de A Coruña

Campus de Elviña s/n, 15071, A Coruña.

{diego.fernandez, fidel.casheda, fjnovoa, xacobe.macia, victor.carneiro}@udc.es

Resumen—Con el auge de las nuevas tecnologías cada vez es más complicado motivar e incentivar a los jóvenes en todos los niveles de la educación. Para acrecentar el estímulo y la atención, en los últimos años se ha expandido el empleo de las mecánicas del juego en un ámbito como el de la enseñanza. A este hecho se le conoce con el término gamificación. En este trabajo se aborda la experiencia vivida en la asignatura de Redes del Grado en Ingeniería Informática en la Universidad de A Coruña. En la programación didáctica de dicha asignatura se incluyó el empleo de un concurso de preguntas a realizar en el propio aula de prácticas. Aunque el análisis realizado sobre las calificaciones no revela ninguna mejora significativa al emplear el concurso, la gamificación trae consigo intrínsecamente una serie de beneficios como el fomento de la motivación, las relaciones sociales, la integración, etc.

Palabras Clave—Gamificación, e-learning, aprendizaje práctico, metodología docente

I. INTRODUCCIÓN

Debido a la gran proliferación de las tecnologías de la información y a su constante evolución, existe la necesidad de formar adecuadamente a los futuros graduados en este ámbito. Sin embargo, este hecho no siempre es sencillo debido a falta de motivación, interés, iniciativa, etc., por parte del alumnado. La progresiva introducción de herramientas de aprendizaje basadas en Web en el ámbito universitario ha facilitado la incorporación de nuevas estrategias educativas. En los últimos años, la “gamificación” (uso de elementos del diseño de juegos en contextos ajenos al juego [1]) ha comenzado a implantarse como un medio que permite estimular la curiosidad, la imaginación y la sensibilidad en el aprendizaje [2], que se pierden en la enseñanza tradicional basada en libros de texto y exámenes. Los jugadores habitualmente han de tener ciertas cualidades que se pretende fomentar en el alumno, como persistencia, saber tomar riesgos, atención al detalle, y capacidad para la resolución de problemas [3]. Aunque el juego tiene como finalidad más habitual el entretenimiento, también se usa en otros dominios como la defensa, gestión de emergencias, planificación urbana, política, religión, salud, educación, investigación científica [4]...

En su blog [5], Yu-kai Chou define la gamificación como “el arte de obtener todos los elementos divertidos y adictivos que se encuentran en los juegos, y aplicarlos a las actividades productivas o del mundo real”. El proceso se diseña centrándose en el ser humano, en lugar de puramente en la eficiencia. En el dominio de los juegos se lleva tiempo desarrollando técnicas para enganchar y motivar a sus usuarios.

Así, por ejemplo, los jóvenes son capaces de pasar mucho tiempo delante de su videoconsola a horas intempestivas matando una y otra vez a un mismo monstruo por el mero hecho de subir un nivel a su personaje. Traspasar esto al mundo de la educación implicaría que los alumnos trabajarían en una determinada materia por pura iniciativa, con el mero fin de lograr comprenderla.

La gamificación, de hecho, es una de las tecnologías que aparecen nombradas en el *NMC Horizon Report* [6], un informe que indica el impacto que tendrán las tecnologías más emergentes en los próximos cinco años.

Por otro lado, con la implantación del Espacio Europeo de Educación Superior (EEES) [7], se ha reestructurado la docencia en base a competencias, que precisan de la adquisición de conceptos teóricos y de habilidades prácticas. Para alcanzar los objetivos planteados en la materia de Redes que se imparte en la titulación de Grado en Ingeniería en Informática de la Universidad de A Coruña, al tiempo que se pretende incrementar el grado de interés y la estimulación del aprendizaje, se ha decidido incorporar a la docencia de la materia elementos propios del diseño de juegos, mediante un concurso de preguntas y respuestas. Esta actividad tiene como objetivo incentivar al alumno en la preparación de la asignatura, por medio de un juego en el que compite con un grupo de compañeros. Por tratarse de un juego, ha de poseer una serie de características [8]: objetivos, reglas, sistema de retroalimentación y participación voluntaria. Además, existen ciertos conceptos muy relacionados como pueden ser las clasificaciones, las puntuaciones, etc. El principal punto de separación entre este tipo de juegos y una actividad al uso es el conjunto de reglas que los definen. Mientras en una actividad normal se proporcionan reglas de tal manera que se sigue el camino óptimo para alcanzar los objetivos, en un juego el camino es sub-óptimo, incluso llegando a introducir obstáculos para alcanzar la meta.

La incorporación de elementos del juego en entornos educativos acarrea potenciales beneficios [9], muy relacionados con el proceso de aprendizaje y la motivación del alumno. Una de las principales ventajas de la gamificación es la participación, la lealtad o el compromiso (*engagement*) del jugador. Son de destacar también los efectos emocionales positivos que aparecen asociados al uso de videojuegos como la curiosidad, la satisfacción, la creatividad, etc. [10]. Sin embargo, en determinadas ocasiones, cuando la gamificación es muy acusada o no está bien planteada, puede suponer efectos negativos en

el alumno. Por ejemplo, puede ser contraproducente ofrecerle grandes recompensas [11]. Esto es debido a que el diseño no se ajusta a la finalidad de motivar a los alumnos para que continúen con la asignatura y de ofrecer ayuda a los profesores [4]. De esta manera los alumnos se pueden acostumbrar a que deberían aprender sólo cuando exista una motivación externa.

El presente artículo está organizado del siguiente modo. En la Sección II se realiza una descripción del contexto en el que se ha realizado la puesta en práctica del juego. A continuación, se explica la mecánica del mismo. Posteriormente, se analiza la influencia en las calificaciones de la inclusión del juego en la programación didáctica de la asignatura. En la Sección V se realiza una discusión acerca de los beneficios y de los problemas asociados con el juego presentado. Finalmente se presentan las conclusiones.

II. DESCRIPCIÓN DE LA ASIGNATURA

El trabajo aquí expuesto se ha realizado en la asignatura de Redes en el curso 2013-2014, materia obligatoria perteneciente al segundo cuatrimestre del segundo curso de Grado en Ingeniería Informática de la Universidad de A Coruña. Se trata de una asignatura de 6 créditos de tipo *European Credit Transfer System* (ECTS); por tanto, ya adaptada al EEES. Entre las competencias que abarca esta materia destaca el conocimiento y aplicación de las características, funcionalidades y estructura de los sistemas distribuidos, las redes de computadores e Internet, y diseñar e implementar aplicaciones basadas en ellas.

Los objetivos principales de la asignatura se pueden resumir en los siguientes puntos:

- Conocer las diferencias fundamentales de los sistemas físicos de transmisión.
- Entender la división de las redes en capas de protocolos.
- Comprender el funcionamiento del enrutamiento y los servicios de red.

Para alcanzar dichos objetivos, la asignatura se divide en diversos bloques, que se describen a continuación:

- Introducción. En este primer bloque se induce al alumnado a hacer una reflexión sobre cuál es el conocimiento previo que tiene sobre los principales conceptos de la asignatura. A continuación, se desarrolla la idea de red en general y se explican los términos esenciales asociados a una arquitectura de red, al modelo *Open System Interconnection* (OSI) y a la pila *Transmission Control Protocol* (TCP)/*Internet Protocol* (IP). El resto de los bloques abordarán los distintos niveles de la pila TCP/IP, siguiendo una aproximación de arriba a abajo [12] [13].
- Nivel de aplicación. En el nivel de aplicación se hace referencia a una serie de protocolos ampliamente empleados como son: *Hypertext Transfer Protocol* (HTTP), *Simple Mail Transfer Protocol* (SMTP), *Post Office Protocol 3* (POP3), *Internet Message Access Protocol* (IMAP), *File Transfer Protocol* (FTP) y *Domain Name System* (DNS).
- Nivel de transporte. En este bloque se describen los protocolos TCP y *User Datagram Protocol* (UDP), incluyendo las conexiones TCP, la transferencia fiable de datos y el intercambio de datos TCP, entre otros apartados.

- Nivel de red. El nivel de red comprende aspectos básicos del protocolo IP. Dos de los conceptos más relevantes de este bloque son las subredes y el enrutamiento. Además de ellos, *Internet Control Message Protocol* (ICMP), IP versión 6 (IPv6), *Dynamic Host Configuration Protocol* (DHCP) o *Network Address Translation* (NAT) son también importantes.
- Nivel de enlace. El último bloque está orientado a explicar las características principales del protocolo Ethernet. También se hace hincapié en otros protocolos como *Address Resolution Protocol* (ARP), y tecnologías inalámbricas.

En cuanto a la metodología empleada, la asignatura se distribuye en 60 horas presenciales, que se reparten entre prácticas de laboratorio, seminarios de grupos reducidos y sesiones magistrales. La plataforma virtual de la universidad (actualmente Moodle versión 2.5) sirve como base para la difusión del material necesario. En las prácticas de laboratorio los alumnos deben profundizar en diversos aspectos teóricos de la asignatura. Para lograrlo, realizan tanto desarrollos de prácticas de programación (por ejemplo, implementan un servidor web sencillo), como prácticas empleando herramientas de emulación/simulación de redes [14] (centrándose en los conceptos de subredes y enrutamiento). Las clases de prácticas son de dos horas de duración, alcanzando un total de 20 horas presenciales. La prácticas forman parte de la evaluación continua de la asignatura. En concreto, un 25% de la nota final proviene de las prácticas.

Con respecto a los seminarios, están concebidos para trabajar con grupos reducidos de alumnos para poder profundizar en determinados contenidos de la materia de una forma más personalizada. Al igual que sucedía con las clases prácticas, los seminarios son de dos horas de duración. Sin embargo, el cómputo total de horas presenciales de seminarios es de 10 horas y el porcentaje de la nota final es de un 5%.

Finalmente, las sesiones magistrales, en las que se fomenta la participación del alumnado, se distribuyen en dos clases a la semana de una hora cada una. En total, suponen 30 horas a lo largo de un cuatrimestre.

III. CONCURSO

A. Descripción

El concurso se realiza en clases de prácticas y es de asistencia voluntaria. Los alumnos se dividen en equipos de no más de 4 personas. Cada uno de estos equipos dispone de un portavoz. El objetivo principal es ayudar a los alumnos a comprender mejor los conceptos teóricos explicados en las sesiones magistrales.

Para la realización del concurso se emplea una clase de prácticas completa en las últimas semanas del curso, de tal manera que los alumnos dispongan de un porcentaje elevado del contenido de la asignatura. De dicha clase la primera hora se dedica a la preparación de preguntas: cada grupo elabora al menos diez preguntas tipo test con tres posibles respuestas cada una. Entre las respuestas sólo una de ellas es correcta. Aunque, como se puede apreciar en la Fig. 1, existen otras alternativas, para el formato de dichas preguntas se decide emplear GIFT¹, el cual está pensado especialmente para la

¹https://docs.moodle.org/23/en/GIFT_format.

Área personal / 614G010171314 / Banco de preguntas / Importar

Importar preguntas de un archivo Expandir todo

Formato de archivo

- Blackboard V6+
- Examview
- Formato Aiken
- Formato de palabra ausente
- Formato GIFT
- Formato Learnwise
- Formato Moodle XML
- Formato WebCT
- Respuestas Incrustadas (Cloze)

General

Categoría a donde importar: Valor por defecto para 614G010171314 - Redes - Grao en Enxeñaría Informática

Obtener categoría de archivo Obtener contexto de archivo

Coincidir calificaciones: Error si la calificación no está en la lista

Detenerse si se produce un error: Si

Fig. 1. Importación de preguntas.

edición de este tipo de cuestionarios. La razón principal es que en la Universidad de A Coruña se emplea Moodle como plataforma de *e-learning*, y GIFT ha sido desarrollado dentro de la Comunidad de Moodle. Una vez redactadas las preguntas, éstas son importadas desde la plataforma.

La segunda hora de clase está destinada a la formulación y a la resolución de las preguntas. Cada equipo realiza dos rondas en las que propone una pregunta a cada uno de los equipos restantes. El tiempo máximo para responder a cada pregunta es un minuto. Si se responde fuera de tiempo se considera que la respuesta es errónea. La pregunta planteada ha de ser respondida únicamente por el portavoz del equipo, que podrá debatir la respuesta y consultar cualquier fuente de información. Se emplea Moodle para responder a las preguntas.

Terminada la clase práctica, el alumno dispondrá en la plataforma de *e-learning* de la opción de realizar cuestionarios creados aleatoriamente a partir de todas las preguntas importadas. El alumno podrá efectuar cuestionarios las veces que quiera y tendrá la posibilidad de obtener estadísticas acerca de sus resultados. Además, el profesor podrá visualizar también estadísticas generales, incluyendo el tiempo empleado en la realización del cuestionario, la calificación, las preguntas que conforman cada cuestionario, las respuestas seleccionadas por el alumno, etc.

B. Puntuación

La puntuación viene determinada por el número de aciertos y el número de errores tanto en las respuestas como en la elaboración de las preguntas.

A la hora de formular la pregunta se puede obtener un punto cuando el equipo rival no sabe la respuesta.

A la hora de responder a una pregunta se consigue un punto si el portavoz responde correctamente a la pregunta o si el equipo que responde detecta que la pregunta propuesta no está formulada correctamente.

La nota obtenida está directamente relacionada con la puntuación alcanzada.

Además, como ya se ha comentado en la sección anterior, existe una clasificación general según el número de tests realizados y el número de preguntas acertadas en la plataforma de *e-learning*.

C. Otras iniciativas

Aparte del concurso anteriormente descrito, se le ha propuesto al alumnado otras actividades adicionales. Entre ellas, se puede destacar un concurso de cortometrajes, cuya participación fue voluntaria. En el curso mencionado (2013-2014) la tarea a realizar fue la siguiente: los cortometrajes participantes debían describir, en un máximo de 3 minutos, el funcionamiento de la navegación web. El objetivo era representar cómo se envían los datos al solicitar y descargar una página web, indicando los distintos protocolos que intervienen (HTTP, TCP, IP...) y sus funcionalidades, tanto en origen como en destino. En la evaluación se valoró especialmente la corrección técnica de las ideas representadas.

Como restricción se decidió que los grupos de trabajo no superasen las cinco personas. Lógicamente, los cortometrajes debían ser aptos para todos los públicos.

Como incentivo, se propuso el incremento de hasta un punto de la nota de la asignatura, siempre y cuando la nota obtenida en la prueba teórica alcanzase una nota de 4.

El concurso tuvo muy buena acogida entre los estudiantes, que presentaron ideas muy originales y correctas desde el punto de vista técnico.

IV. ANÁLISIS DE RESULTADOS

La primera cuestión que se plantea es investigar si existe una relación directa entre la nota que ha obtenido un alumno en el concurso y la nota que obtiene en el examen de la asignatura de Redes. Para ello se analizan las notas del resultado del concurso correspondientes al curso 2013-2014 y las notas obtenidas en el examen de la convocatoria de junio.

En total 161 alumnos participaron en el concurso obteniendo una nota media de 7,7, mientras que al examen de junio se presentaron un total de 172 alumnos con una nota media de 4,7 y una desviación típica de 1,9. En la Fig. 2 se muestra el diagrama de dispersión de puntos de la nota obtenida en el curso, frente a la nota obtenida en el examen de la asignatura en la convocatoria de junio.

Como se puede observar en la gráfica, no existe una relación clara entre la nota obtenida en el concurso y la nota obtenida en el examen. Al realizar un análisis de regresión entre ambas variables se confirma el bajo nivel explicativo de la nota del concurso sobre la nota del examen con un coeficiente de determinación bajo ($R^2 = 0,136$). Comentar que el análisis obtiene un coeficiente de regresión de 0,179, lo que indica una relación positiva y directa entre ambas variables, y que se puede observar en la figura como una ligera tendencia a obtener notas más elevadas por aquellos alumnos que han sacado una nota elevada en el concurso.

La segunda cuestión que se ha analizado es si la realización del concurso permite mejorar la tasa de aprobados de la asignatura. En este sentido, se han analizado las notas obtenidas en el examen de junio por los alumnos del curso 2012-2013, donde no se ha realizado el ejercicio y las notas obtenidas por

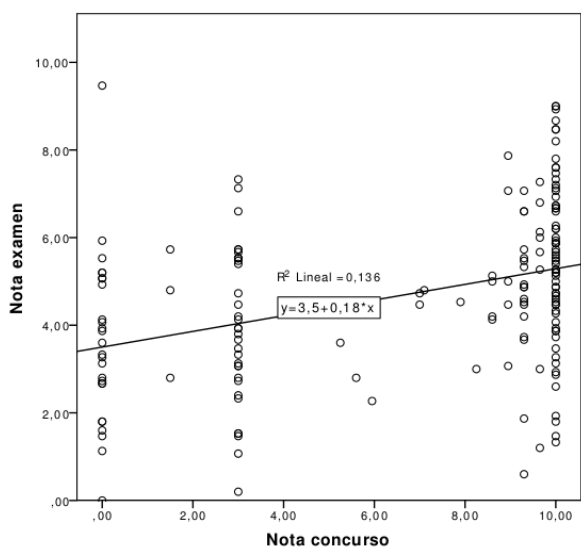


Fig. 2. Gráfico de dispersión.

Tabla I
TABLA DE CONTINGENCIA COMPLETA.

	No Presentados	Suspensos	Aprobados	Total
Curso 2012 – 2013	57	89	86	232
Curso 2013 – 2014	43	94	78	215
Total	100	183	164	

los alumnos del curso 2013-2014, en donde sí han participado en el concurso. En cada caso se han valorado los alumnos que no se han presentado al examen (etiquetados como NP), los que han suspendido y los que han aprobado. La Tabla I muestra la tabla de contingencia para el análisis realizado.

En el curso 2012-2013 había un total de 232 alumnos matriculados, de los cuales 86 aprobaron el examen de junio (la nota media obtenida fue de 4,8 con una desviación típica de 1,9). En el curso 2013-2014 se matricularon 215 alumnos aprobando 78 con una nota media de 4,7 en el examen de junio. El análisis Chi-cuadrado sobre estos datos permite comprobar si existe alguna diferencia entre ambos cursos, y en este caso indica que se debe aceptar la hipótesis nula (las notas obtenidas en los dos cursos son iguales), con un p-valor de 0,40.

Se realiza una segunda variante del análisis, teniendo en cuenta únicamente aquellos alumnos que se han presentado al examen de la asignatura, obteniendo una tabla de contingencia similar a la Tabla I pero sin la columna con el número de alumnos no presentados.

En este caso, al igual que en el caso anterior, el análisis de Chi-cuadrado lleva a aceptar que las notas de ambos cursos no se diferencian, con un p-valor de 0,48.

V. DISCUSIÓN

Como se ha comprobado en la sección IV, los análisis realizados determinan que no existen diferencias entre un curso en el que se realizó el concurso y otro en el que no se propuso. Profundizando en la cuestión, podemos encontrar diversas razones que justifiquen este resultado. En primer

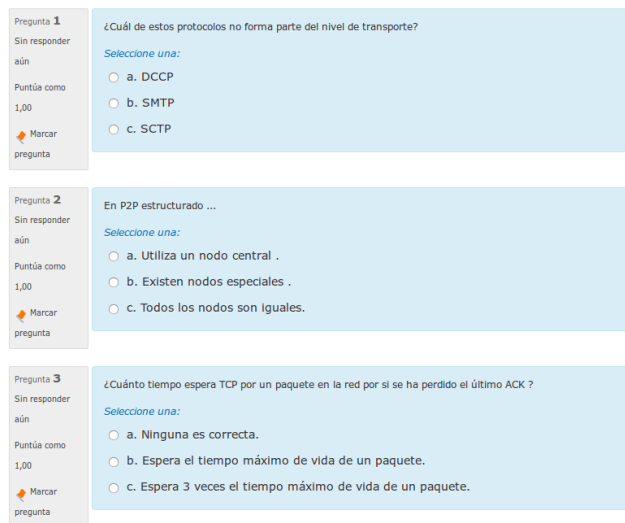


Fig. 3. Vista de un cuestionario en Moodle con un rol de alumno.

lugar, el nivel de cada promoción de alumnos puede variar. Para poder diseñar mejor el experimento, habría que haber hecho una selección aleatoria de los alumnos de un mismo curso, y separarlos en grupos que realizasen el concurso y grupos que no lo hiciesen. Esto, en su momento, no se contempló como alternativa plausible, puesto que se intentaba obtener la crítica de los alumnos y saber cuál era su reacción a la propuesta.

Por otro lado, los alumnos, para poder realizar el concurso, asisten en diferentes horarios al seminario en cuestión. De hecho, hay 12 grupos diferentes que no tienen por qué ser homogéneos, ni en número, ni en el nivel de los estudiantes. La evaluación del seminario se realizó individualmente dentro de cada grupo. De este modo, si en un mismo grupo, por azar, coincidían los alumnos más brillantes de la promoción, podría resultar que esos mismos alumnos se viesen perjudicados en la evaluación obtenida. Esta variabilidad no está explicada en el estudio realizado.

Aunque el estudio no es concluyente, a continuación se proporcionan un conjunto de argumentos que apoyan la elección del concurso como una actividad dentro de la asignatura:

- Proporciona una gran colección de preguntas. Los propios alumnos elaboran un conjunto de preguntas que el profesor de su grupo corrige y posteriormente pone a disposición de todos los alumnos de la asignatura. De esta manera, pueden autenticarse en la plataforma de *e-learning* y realizar un número elevado de tests diferentes propuestos por ellos mismos. Cada test estará compuesto de un subconjunto de preguntas seleccionadas aleatoriamente a partir de todas las preguntas disponibles. En la Fig. 3 se muestra parte de un test de ejemplo tal y como lo vería un alumno. El alumno también dispone de un tiempo reducido para responder al test (10 minutos).
- Aporta retroalimentación tanto para alumnos como para profesores. Como se ha comentado en la Sección I, una de las características que ha de poseer el concurso, por ser un juego, es la de disponer de un sistema de retroalimentación. Así, conforme el alumno va intentando solucionar

cuestionarios, recibe retroalimentación por parte de la plataforma, para que pueda adoptar una actitud crítica del grado de conocimiento que tiene acerca del contenido de la asignatura. De este modo, además, el alumno se motiva a medida que sus resultados van mejorando.

Según el número de cuestionarios resueltos por los alumnos se incrementa, el profesor también dispone de más información y estadísticas acerca del número de intentos de cada alumno, número de preguntas falladas en los intentos, preguntas más difíciles de las mostradas, etc. De esta forma recibe una información valiosa acerca de aquellos temas que menos se han comprendido en un momento en el que todavía está a tiempo para ofrecer alguna hora de refuerzo. De otro modo, sería en el examen teórico de la asignatura cuando recibiría la información, momento en el cual ya no habría forma de proponer horas para explicar lo que fuese necesario.

- Induce al alumno a tener que preparar la asignatura con cierta antelación. Debido al reducido periodo de tiempo del que disponen para contestar a cada pregunta, los alumnos deben poder analizar las alternativas rápidamente, sintetizar los enunciados (algunos de ellos especialmente largos), y desarrollar un pensamiento crítico para poder así proporcionar una respuesta e, incluso, encontrar posibles errores no detectados por el profesor en los enunciados de sus rivales. Para ello es necesario un estudio previo del contenido de la asignatura.

A pesar de que el contenido de la asignatura no es demasiado denso, sí es lo suficiente para que generalmente el alumno necesite un tiempo para poder asentar determinados conceptos. Al ir recorriendo los diferentes niveles de la pila TCP/IP, no siempre es sencillo comprender las relaciones existentes entre ellos, cómo se complementan. Es por esa razón que es de gran ayuda una actividad como la que se está analizando.

Asimismo, el preparar previamente el contenido del concurso puede acarrear que el alumno retenga mejor la solución a las respuestas (tanto acertadas, como, sobre todo, erradas), de tal modo que puede después basarse en dicha experiencia para resolver problemas similares.

- Cubre competencias transversales y nucleares del título. Los planes de estudios de los grados en Ingeniería no suelen tener en consideración asignaturas cuyo principal objetivo sea poder hacer frente a una entrevista, una exposición, etc. Si bien el concurso dista mucho de ser una situación equivalente, sí es cierto que es mucho más propenso a la participación general del alumnado, y no de un grupo reducido, fomentando así la integración y la socialización. Como resultado, en muchas ocasiones, alumnos con problemas en sus relaciones sociales intentan preparar la asignatura para conseguir su integración en otros grupos. Incluso, alumnos con mucha aceptación social, intentan también prepararse por miedo a demostrar su falta de conocimiento acerca de la materia.

Este concurso, por tanto, sirve de vía para trabajar las expresiones oral y escrita del alumno, utilizando la terminología de la asignatura de forma adecuada.

Además, permite incidir en la relevancia de una actitud crítica (para poder detectar errores en los enunciados, por ejemplo), otra competencia transversal importante.

- Promueve el estudio por medio de la competitividad. La competitividad es un valor por sí mismo que puede motivar a muchos alumnos a participar en el concurso y preparar la asignatura.

Sin embargo, en el curso 2012-2013 se había propuesto también el concurso, pero la participación no fue demasiado elevada. La razón es simple: no disponían de un incentivo en la calificación que los motivase a participar. El premio por participar era la visita a un Centro de Procesamiento de Datos (CPD) de una empresa gallega. En el curso 2013-2014, en donde la posición en el concurso se reflejó tímidamente en la calificación final del alumno, la participación se incrementó notablemente. Además, se apreció también una evolución en el comportamiento de los alumnos conforme transcurrían los días del concurso: el factor sorpresa ante la novedad de la mecánica del concurso provocó que los primeros días los grupos fuesen menos preparados.

- Estimula la creatividad y la imaginación. El disponer de medios telemáticos para poder resolver las preguntas incita al alumno a hacer uso de la creatividad y de la imaginación para poder desarrollar enunciados no obvios a partir del contenido de la asignatura.

- Permite entrenar la gestión de equipos. La colaboración, el fomento de la discusión y la toma de iniciativa [15] son aspectos que toman especial relevancia en este apartado y que posteriormente en su vida laboral serán también de vital importancia para la mayor parte de los alumnos. Analizando el conjunto de los equipos, resulta muy interesante percibir cómo el comportamiento de los grupos difiere en gran medida. Debido a que el número de errores y aciertos en las preguntas traía en consecuencia una reducción o un incremento en la nota de los componentes del propio equipo y de los equipos competidores, se apreciaron dos tendencias completamente opuestas en muchos de los grupos: mientras que algunos de ellos miraban por el bien común, planteando preguntas sumamente sencillas que todo el mundo pudiese responder correctamente, otros presentaron una conducta muy agresiva, enviando preguntas muy complejas a sus rivales, buscando únicamente el beneficio del propio equipo. Este hecho originó situaciones de descompensación; es decir, grupos de prácticas en los que todos los alumnos alcanzaban buenas notas, y grupos en los que existía una gran variabilidad en las mismas.

Por otro lado, también es sumamente interesante apreciar el rol de cada alumno dentro de su equipo. Así, existen equipos que son dirigidos claramente por un único alumno, mientras en otros todos los componentes del equipo están a un mismo nivel.

- Es divertido. Por el hecho de tratarse de un concurso (un juego, al fin y al cabo), los alumnos generalmente se encuentran más relajados. Esto trae como consecuencia que se produzcan momentos divertidos, como, por ejemplo, la elección del nombre del grupo, y se reduzca el aburrimiento (demasiado frecuente en sesiones

magistrales). De esta manera, se facilita la comunicación entre los alumnos (reduciendo la timidez), y se descargan tensiones, potenciando así la creatividad y la imaginación, como ya se ha comentado.

En contraposición a estos beneficios, existen también una serie de problemas que pueden surgir en torno a la gamificación y que deben de ser tenidos en cuenta [16]. Entre ellos figuran los siguientes:

- Diseñar erróneamente el juego. Si bien la estructura del juego planteado en este artículo no es compleja, el diseño del juego ha de adaptarse al tipo de usuario al que está dirigido. De no ser así, el objetivo probablemente no se cumpla, puesto que los alumnos no se motivarían lo suficiente para participar en él.
- No disponer de experiencia con la tecnología. Aunque en la época actual proliferan las tecnologías de la información y las comunicaciones, los alumnos no tienen por qué disponer de conocimiento acerca del entorno. En el caso de la asignatura descrita en la Sección II, los alumnos sí poseen un conocimiento elevado, por lo que no manifiestan dicha dificultad. Pero no sólo es conveniente que tengan experiencia los alumnos, sino también los profesores. Para hacer una integración completa del juego en una programación didáctica, es necesario un cambio en la metodología, en la relación alumno-profesor, en la evaluación, etc.
- No tener claros los objetivos. Es necesario adaptar los objetivos de aprendizaje para poder ajustarse al juego, en lugar de adaptar el juego para poder ajustarse a los objetivos de aprendizaje. De esta manera, el profesor debería volver a diseñar sus lecciones a partir de juegos que considere útiles para ser utilizados por los alumnos.
- No poder adaptar el tiempo de los juegos disponibles al número de horas de prácticas. El concurso descrito en este artículo se ajusta a la duración de una clase práctica (dos horas). Sin embargo, es notable la falta de juegos relevantes desde el punto de vista educacional que se ajusten a un número de horas reducido para poder ser empleados en el aula.

Finalmente, de un modo informal se recopiló información acerca de la impresión de los alumnos acerca del concurso. En general, las respuestas fueron muy positivas. Sin embargo, también hubo una reclamación anónima por parte de un equipo, ya que éste consideraba que se había visto perjudicado en el concurso, porque el número de integrantes de todos los equipos no era homogéneo.

VI. CONCLUSIONES

Este artículo describe la experiencia vivida en un curso de Redes en un Grado de Ingeniería Informática al proponer como actividad didáctica un concurso de preguntas y respuestas. En él los alumnos se agrupan por equipos, plantean preguntas a otros compañeros empleando la plataforma corporativa de *e-learning*, e intentan resolver las preguntas que les realizan. El objetivo principal era ayudar a los estudiantes a asimilar los conceptos teóricos explicados en las sesiones magistrales de un modo más entretenido para ellos.

Introducir la mecánica de los juegos en la asignatura ha proporcionado importantes beneficios, principalmente en cuanto a

la motivación del alumnado a la hora de preparar la asignatura y afrontar el estudio de los aspectos teóricos de la misma. Además, también se estimulan otros aspectos emocionales como la curiosidad (en el momento de profundizar en un determinado concepto), la creatividad (a la hora de desarrollar preguntas no triviales que otros equipos no puedan deducir con el simple hecho de leer apuntes de la asignatura), la satisfacción personal (cuando el equipo compite con éxito en el juego), etc.

Desde el punto de vista del docente, los resultados del concurso proporcionan información acerca de los conceptos que han quedado menos claros, aquéllos que plantean más dudas, etc.

En el futuro, dado que la satisfacción, el orgullo personal tras haber respondido correctamente a las preguntas, el reconocimiento o la motivación son aspectos deseables en el alumnado, se buscarán nuevas alternativas al concurso planteado. Por ejemplo, se podrían plantear pequeños juegos con el contenido de la asignatura, de tal manera que los alumnos puedan ir profundizando individualmente en cada tema a lo largo de todo el cuatrimestre.

AGRADECIMIENTOS

Este trabajo ha sido financiado por la Consellería de Cultura, Educación e Ordenación Universitaria con su ayuda GPC2014/045 (Grupos con Potencial de Crecimiento).

Los autores quieren agradecer también la aportación de los profesores del Grupo de Telemática de la Universidad de A Coruña, así como la participación y experimentación en el proyecto del personal de la asignatura de Redes.

REFERENCIAS

- [1] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification";" in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, (New York, NY, USA), pp. 9–15, ACM, 2011.
- [2] D. Thomas and J. S. Brown, *A new culture of learning : cultivating the imagination for a world of constant change*, vol. 219. Lexington, Ky.: CreateSpace, 2011.
- [3] E. Klopfer, S. Osterweil, and K. Salen, "Moving learning games forward: Obstacles, opportunities and openness." The Education Arcade. Massachusetts Institute of Technology, 2009.
- [4] C. I. Muntean, "Raising engagement in e-learning through gamification;" in *Proc. 6th International Conference on Virtual Learning ICVL*, pp. 323–329, 2011.
- [5] Y. Chou, "What is gamification.;" [online] <http://www.yukaichou.com/gamification-examples/what-is-gamification/>, 2012.
- [6] L. Johnson, S. Adams, M. Cummins, V. Estrada, A. Freeman, and H. Ludgate, "The NMC Horizon Report: 2013 Higher Education Edition," tech. rep., New Media Consortium.
- [7] European Commission/EACEA/Eurydice, "The European Higher Education Area in 2015: Bologna Process Implementation Report." [online] http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/182EN.pdf, 2015.
- [8] P. Callaghan, "Games and Activities. Gamification, Education, and Ideas." [online] <http://peardarcallaghan.com/gamification/games-and-activities/>, 2015.
- [9] J. J. Lee and J. Hammer, "Gamification in education: What, how, why bother?;" *Academic Exchange Quarterly*, vol. 15, no. 2, p. 2, 2011.
- [10] J. McGonigal, *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. The Penguin Group, 2011.
- [11] M. D. Hanus and J. Fox, "Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance;" *Computers & Education*, vol. 80, pp. 152–161, 2015.
- [12] J. Kurose, K. Ross, and M. Barrio Solórzano, *Redes de computadoras: Un enfoque descendente basado en Internet*. Pearson Educación, 2004.

- [13] W. R. Stevens in *TCP/IP Illustrated (Vol. 1): The Protocols*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1993.
- [14] D. Fernandez, F. Cacheda, V. Formoso, and V. Carneiro, "Experiences on the laboratory of a computer networking course," *ICERI2012 Proceedings*, pp. 3868–3877, 2012.
- [15] M. Borys and M. Laskowski, "Implementing Game Elements into Didactic Process: A Case Study," in *Active Citizenship by Knowledge Management & Innovation: Proceedings of the Management, Knowledge and Learning International Conference 2013*, pp. 819–824, ToKnowPress, ToKnowPress, 2013.
- [16] J. Marquis, "The trouble with gamification.." [online] <http://www.onlineuniversities.com/blog/2012/07/the-trouble-gamification/>, 2012.

