JITEL 2021

LIBRO DE ACTAS

XV Jornadas de Ingeniería Telemática

A CORUÑA 2021

ISBN: 978-84-09-35131-2

Editores:
Victor Manuel Carneiro Díaz
Laura Victoria Vigoya Morales

XV Jornadas de Ingeniería Telemática – JITEL 2021

JITEL 2021

Presentación

Una de las lecciones aprendidas, en este último año y medio de pandemia COVID-19, es que los avances y resultados, en el ámbito de la Telemática, han sido decisivos para mantener la actividad en múltiples aspectos de la sociedad. El teletrabajo, la teleasistencia, la teledocencia y los sistemas de comunicación personal a distancia, no sólo han aliviado la complicada situación económica que ha generado la pandemia, sino que han contribuido a generar una nueva cultura de relaciones sociales y empresariales que ha llegado para quedarse. 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 decimoquinta 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.

La edición de 2021, celebrada en A Coruña, ha sido organizada por el área de Ingeniería Telemática de la Universidad de A Coruña, el Centro de Innovación en Tecnologías de la Información y las Comunicaciones (CITIC) y la Sociedad Científica de Ingeniería Telemática (SCITEL).

Estas actas contienen las 58 contribuciones aceptadas en las jornadas, cuyas sesiones se organizan en los siguientes ámbitos temáticos: Docencia en telemática; Multimedia, salud y sociedad digitales; Análisis de datos de redes (machine learning), cloud and fog computing; Aplicaciones y servicios; Seguridad en comunicaciones, redes y sistemas. Tecnología Blockchain; Virtualización de redes y servicios (SDN/NFV, orquestación de recursos, slicing…); Gestión y operación de redes y sistemas; Redes de nueva generación (5G, 6G, …); Redes dedicadas (IoT, m2m, e2e, redes de sensores, redes Ad-Hoc…); Calidad en comunicaciones, redes y sistemas (parámetros y percepción); además de una sesión de presentación de líneas de investigación.

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

Víctor Manuel Carneiro Díaz
Comité de Programa de JITEL 2021

JITEL 2021

Comité de Programa

Presidente:
Victor Carneiro Díaz (Universidade de A Coruña)

Miembros del Comité de Programa:
Álvaro Suárez Sarmiento (Universidad de Las Palmas de Gran Canaria)
Maria Victoria Higuero Aperribay (Euskal Herriko Unibertsitatea)
Luis De La Cruz Piris (Univesidad de Alcalá)
Ramón Agüero Calvo (Universidad de Cantabria)
Iria Manuela Estevez Ayres (Universidad Carlos III de Madrid)
Javier Carmona Murillo (Universidad de Extremadura)
Jorge Navarro Ortiz (Universidad de Granada)
Magdalena Payeras Capellà (Universitat Illes Balears)
Sebastián García Galán (Universidad de Jaén)
Rodrigo Román Castro (Universidad de Málaga)
Guiomar Corral Torruella (Universitat Ramon Llull)
Antonio Ruiz Martínez (Universidad de Murcia)
Roberto García Fernández (Universidad de Oviedo)
Miguel Ángel Valero Duboy (Universidad Politécnica de Madrid)
Pilar Manzanares López (Universidad Politécnica de Cartagena)
Xavier Hesselbach Serra (Universidad Politècnica de Catalunya)
Jaime Lloret Mauri (Universitat Politècnica de Valencia)
Antonio José Estepa Alonso (Universidad de Sevilla)
Santiago Felici Castell (Universitat de Valencia)
Miguel Luis Bote Lorenzo (Universidad de Valladolid)
Julián Fernández Navajas (Universidad de Zaragoza)

Comité de Organización

Fidel Cacheda Seijo (Universidade de A Coruña)
Comité de Dirección del CITIC
Laura Vigoya Morales (Universidade de A Coruña)
María Jesús Vidal Insua (CITIC)
Alejandro Mosteiro Vázquez (CITIC)
Diego Fernández Iglesias (Universidade de A Coruña)
Francisco Javier Novoa De Manuel (Universidade de A Coruña)
Manuel Fernández López-Vizcaino (Universidade de A Coruña)

JITEL 2021

Revisores

Aguado, Marina (Euskal Herriko Unibertsitatea)
Agudo, Isaac (Universidad de Málaga)
Agüero Calvo, Ramón (Universidad de Cantabria)
Alario-Hoyos, Carlos (Universidad Carlos III de Madrid)
Alcaraz, Cristina (Universidad de Málaga)
Arco, José (Universidad de Alcalá de Henares)
Asensio-Pérez, Juan (Universidad de Valladolid)
Atutxa, Asier (Euskal Herriko Unibertsitatea)
Azuara, Guillermo (Universidad de Zaragoza)
Barba, Antonio (Universitat Politècnica de Catalunya)
Bernal Bernabe, Jorge (Universidad de Murcia)
Cabot, Miquel A. (Universitat de les Illes Balears)
Cacheda, Fidel (Universidade da Coruña)
Calderon Pastor, María (University Carlos III of Madrid)
Calle-Cancho, Jesús (Universidad de Extremadura)
Canales, María (Universidad de Zaragoza)
Carmona Murillo, Javier (Universidad de Extremadura)
Carral, Juan-Antonio (Universidad de Alcalá de Henares)
Chinchilla-Romero, Natalia (Universidad de Granada)
Corcoba Magaña, Víctor (Universidad de Oviedo)
Corral Torruella, Guiomar (Universitat Ramón Llull)
Cortés-Polo, David (Universidad de Extremadura)
Cruz-Piris, Luis (Universidad de Alcalá)
de la Cruz, Luis (Universitat Politècnica de Catalunya)
Delgado-Ferro, Félix (Universidad de Granada)
Diez, Luis (Universidad de Cantabria)
Dimitriadis, Yannis (Universidad de Valladolid)
Estepa, Antonio (Universidad de Sevilla)
Felici Castell, Santiago (Universidad de Valencia)
Fernández Iglesias, Diego (Universidad de A Coruña)
Fernández-Navajas, Julián (Universidad de Zaragoza)
Fisteus, Jesús (Universidad Carlos III de Madrid)
Franco, David (Euskal Herriko Unibertsitatea)
Gallego-Madrid, Jorge (Universidad de Murcia)
García, Antonio (Universidad de Alcalá de Henares)
García, Marta (Universidad de Cantabria)
García Rubio, Carlos (Universidad Carlos III de Madrid)
García-Carrillo, Dan (Universidad de Oviedo)
García-Pineda, Miguel (Universidad de Valencia)
Giménez Guzmán, José Manuel (Universidad de Alcalá)
Gómez Sánchez, Eduardo (Universidad de Valladolid)
Hesselbach Serra, Xavier (Universitat Politècnica de Catalunya)
Higuero Aperribay, Marivi (Euskal Herriko Unibertsitatea)
Huguet, Llorenç (Universitat de les Illes Balears)
Ibañez, María (Universidad Carlos III de Madrid)

Irastorza, José Angel (Universidad de Cantabria)
Lanza, Jorge (Universidad de Cantabria)
Larrabeiti López, David (Universidad Carlos III de Madrid)
Lloret, Jaime (Universitat Politécnica de Valencia)
López-Vizcaíno, Manuel (Universidade da Coruña)
Macias Lopez, Elsa (Universidad Las Palmas de Gran Canaria)
Malgosa-Sanahuja, José María (Universidad Politécnica de Cartagena)
Manzanares López, Pilar (Universidad Politécnica de Cartagena)
Marrero, Domingo (Universidad de Las Palmas de Gran Canaria)
Martínez, Juan Antonio (Universidad de Murcia)
Matheu, Sara (Universidad de Murcia)
Mayor, Vicente (Universidad de Sevilla)
Melendi, David (Universidad de Oviedo)
Montagud, Mario (Fundació i2CAT, Universitat de València)
Muñoz-Calle, Javier (Universidad de Sevilla)
Muñoz, Antonio (Universidad de Málaga)
Muñoz-Gea, Juan Pedro (Universidad Politécnica de Cartagena)
Mut Puigserver, Macià (Universitat de les Illes Balears)
Navarro Ortiz, Jorge (Universidad de Granada)
Novoa De Manuel, Francisco Javier (Universidade da Coruña)
Payeras-Capellà, M. Magdalena (Universitat de les Illes Balears)
Pozueco, Laura (Universidad de Oviedo)
Riera-Palou, Felip (University of Balearic Islands)
Rios, Ruben (Universidad de Málaga)
Rodríguez-Pérez, Francisco-Javier (Universidad de Extremadura)
Rojas Sánchez, Elisa (Universidad de Alcalá)
Roman Castro, Rodrigo (Universidad de Málaga)
Ruiz-Mas, José (Universidad de Zaragoza)
Samper, José Javier (Universitat de Valencia)
Sanchez, Luis (Universidad de Cantabria)
Sanchez-Iborra, Ramon (Universidad de Murcia)
Sanz Rekalde, Ane (Euskal Herriko Unibertsitatea)
Sasiain, Jorge (Euskal Herriko Unibertsitatea)
Segura García, Jaume (Universitat de Valencia)
Serrat, Joan (Universitat Politècnica de Catalunya)
Soriano Asensi, Antonio (Universitat de Valencia)
Suárez, Álvaro (Universidad Las Palmas de Gran Canaria)
Valera, Francisco (Universidad Carlos III de Madrid)

TABLA DE CONTENIDOS

Redes dedicadas (I)

Redes dedicadas (II)

Redes dedicadas (III) y líneas de investigación

Docencia en Telemática

Multimedia, salud y sociedad digital

Aplicaciones y servicios

Virtualización de redes y servicios. Blockchain.

Seguridad en Comunicaciones, redes y sistemas (I)

Seguridad en Comunicaciones, redes y sistemas (II)

Redes de nueva generación

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Evitando el efecto *random walk* en osciladores de bajo coste para mejorar la sincronización entre nodos WSN

S. Felici-Castell*, JJ. Perez-Solano*, A. Soriano-Asensi*, J. Lopez-Ballester*, J. Segura-Garcia*, E. A. Vargas,† G. Mas‡
* ETSE, Universitat de València,† DEI, Universidad Católica Asunción (PY),‡ BSG Ingenieros
felici@uv.es, jjperezs@uv.es, soan@uv.es, jesus.lopez-ballester@uv.es, jsegura@uv.es, evargas@uc.edu.py, gemma@bsg.es

Resumen

Las redes de sensores inalámbricos (Wireless Sensor Nodes) están en continua evolución y mejorando sus prestaciones introduciendo nuevas tecnologías. Para poder desplegarse a gran escala, estas redes se basan en nodos inalámbricos de bajo coste. Los componentes de bajo coste, y en particular los sistemas de reloj de estos nodos, les impiden lograr una sincronización precisa, que es importante para reducir el número de paquetes transmitidos así como necesaria para ciertas aplicaciones (localización de eventos sonoros, beamforming colaborativo, etc.). Para mejorar la sincronización entre nodos, en este artículo proponemos el uso de un nodo inalámbrico basado en tecnología Ultra Wide Band utilizando componentes disponibles en el mercado, como el transceptor DWM1000 de Decawave. Para ello, modelamos el comportamiento de los relojes utilizando paseos aleatorios (random walks) y proponemos diferentes alternativas para conseguir una sincronización temporal a escala de subnanosegundos utilizando regresiones, con un protocolo de sincronización *one way*. Se concluye que con estos nodos se puede llegar a conseguir un error de sincronización temporal de 149 ps entre transceptores.

***Palabras Clave*—redes dedicadas, ultra wide band, wireless sensor networks, time synchronization, random walk**

I. INTRODUCCIÓN

Las Redes de Sensores Inalámbricos (Wireless Sensor Nodes, WSN) se enfrentan a nuevos retos y aplicaciones complejas como la vigilancia, la localización de eventos sonoros, la formación de haces colaborativos [1] por nombrar algunos. Los despliegues de estas redes suelen tener requisitos de bajo coste y bajo consumo de energía. Estas redes se basan en nodos inalámbricos colaborativos que tienen sus propias unidades de procesamiento, sensores, memorias y comunicaciones inalámbricas, basadas en componentes de bajo coste para abaratarlas.

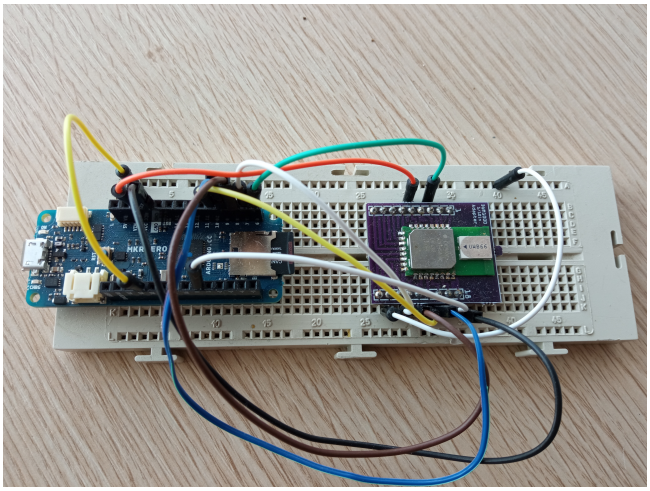


Figura 1: Foto del prototipo del nodo inalámbrico basado en el transceptor Decawave DWM1000 UWB conectado a Arduino MKRZero

En este contexto, el estándar IEEE802.15.4 centrado en capas físicas para velocidades binarias bajas y de bajo consumo, propone las tecnologías Ultra Wide Band (UWB) como una de las alternativas a considerar en WSN. Nos centraremos en esta tecnología, conocida como 'pulse radio', ya que se basa en la transmisión de pulsos de corta duración que requieren de un gran ancho de banda. La tecnología UWB permite realizar mediciones precisas del tiempo de vuelo (Time of Fligth, TOF) en entornos densos con múltiples trayectorias.

Existen transceptores UWB comerciales, como el DWM1000 [2] de Decawave, que cumplen con la norma IEEE802.15.4 (en particular IEEE802.15.4-2011) y que pueden conectarse a un microcontrolador (MCU) como se muestra en la Fig. 1, cumpliendo los requisitos men-

cionados anteriormente (bajo coste y bajo consumo). Este nodo inalámbrico es un prototipo utilizado y testeado en el presente trabajo.

Este transceptor UWB utiliza el marcado de tiempo (time stamping) basado en el mecanismo conocido como Delimitador de Inicio de Trama (Start Frame Delimiter, SFD), permitiendo introducir la marca de tiempos en las tramas cuando se detecta este campo SFD, tanto en las tramas entrantes como en las salientes. Esta característica hace muy interesante la medición del TOF en aplicaciones de localización, para las que fue diseñado este transceptor [3]. Aprovechando la sinergia entre las aplicaciones de localización y la sincronización temporal. Nuestro objetivo es explotar las interesantes características de este transceptor desde un punto de vista diferente, para un uso distinto al de las aplicaciones de localización.

El mecanismo de marcado de tiempo SFD reduce los retrasos no deterministas introducidos por el proceso de comunicación, en particular por la capa de control de acceso al medio (Media Access Control, MAC). Este mecanismo también se ha explotado en otros protocolos de sincronización utilizados en las WSN [4] [5] [6]. Además, estos protocolos se basan en la sincronización temporal a largo plazo, utilizando algoritmos de regresión lineal sobre las últimas marcas de tiempo observadas.

Sin embargo, los relojes de estos nodos en general sufren derivas de reloj, produciendo desviaciones de reloj entre ellos. La desviación entre dos relojes (o skew) se define como la tasa de cambio entre ellos. Estas derivas se deben a su bajas prestaciones, y en particular se ven afectadas por las condiciones ambientales (como la temperatura, la humedad, el voltaje, el envejecimiento, etc.)[7][8], con valores típicos en el rango de 1 parte por millón (ppm) a 100 ppm [9]. Se puede considerar que el skew del reloj varía dinámicamente siguiendo una distribución de ruido blanco gaussiano, cuyo efecto se conoce como Random Walk (RW) [10] [11]. En este trabajo se analiza este comportamiento y tras revisar los trabajos relacionados, se plantean diferentes alternativas para mejorar la sincronización temporal en el diseño propuesto (Fig. 1), con el fin de conseguir una sincronización por debajo de nanosegundos, evitando en lo posible el efecto RW.

El resto del trabajo está estructurado como sigue. En la sección II, mostramos el trabajo relacionado. En la sección III, analizamos el efecto del paseo aleatorio en la desviación del reloj entre los nodos de la WSN. En la Sección IV, introducimos un sencillo protocolo de sincronización. En la Sección V, describimos la propuesta de nodo inalámbrico basado en UWB. En la Sección VI, presentamos los resultados y realizamos un análisis exhaustivo de las diferentes marcas de tiempo utilizando la característica SFD. Finalmente, en la Sección VII, resumimos las principales conclusiones del trabajo.

II. ESTADO DEL ARTE

La sincronización es un tema candente en las WSN. Cabe mencionar protocolos de sincronización como Timing-sync Protocol for Sensor Networks (TPSN) [12], Flooding Time-Synch Protocol (FTSP) [5], Rate Adaptive Time Synchronization (RATS) [4] por nombrar algunos. Estas referencias logran una precisión de sincronización de tiempo del orden de microsegundos, utilizando motas tradicionales como TelosB [13].

Sin embargo, utilizando la tecnología UWB, podemos mejorar esta precisión de sincronización temporal. En [14] se utiliza el kit Decawave DWM1001 Real-Time Location System (DRTLS) (basado en el transceptor DWM1000 UWB), con el que sus autores consiguen generar pulsos de referencia temporal con una fluctuación máxima de 3,3 μs y una desviación estándar de 0,7 μs sin necesidad de hardware adicional. En [15], los autores diseñan un nodo inalámbrico basado en el mismo transceptor, pero añadiendo un circuito de bucle de enganche en fase (Phase Lock Loop, PLL) basado en un reloj atómico a escala de chip, que puede lograr una sincronización de reloj mejor que 5 ns entre dispositivos de interior o sin GPS, con una media de 2,12 ns (desviación estándar de 0,84 ns). Aunque este trabajo proporciona resultados muy interesantes, hay que destacar que en realidad el subsistema de reloj (reloj atómico a escala de chip) es caro, no siendo una solución de bajo coste factible en la práctica.

Por último, analizando el efecto RW en los relojes, si un temporizador se deja sin corregir, su crecimiento de error de temporización puede ser modelado por un RW unidimensional [16] o un proceso de Wiener. En [17] se describe un protocolo para estimar el skew y tiempo de vuelo, utilizando un modelo cuadrático para modelar el RW. Basándose en sus mediciones, su modelo es más preciso que un modelo lineal, concluyendo que el uso de un modelo cuadrático permite incrementar la validez temporal y que el modelo lineal puede funcionar sólo para períodos de tiempo cortos.

De todas estas referencias se desprende que la UWB puede proporcionar mejores resultados en la sincronización de las WSN, pero falta la evaluación del rendimiento de esta tecnología aplicada a la sincronización, y en particular cuando se utilizan productos disponibles en el mercado en despliegues reales.

III. EFECTO "RANDOM WALK" EN OSCILADORES DE BAJO COSTES

Para modelar el efecto RW, introduciremos el modelo general del tiempo entre dos relojes. Para ello, podemos asumir un comportamiento lineal para modelar la relación entre el tiempo local de dos nodos como $y(t) = m{\cdot}x(t){+}h$ donde $y(t)$ y $x(t)$ son sus relojes locales, t es el tiempo global, m es la pendiente (por defecto 1), y h es el desfase del reloj en $t = 0$.

Dado que las condiciones ambientales cambian gradualmente, el efecto de RW se introduce sobre la pendiente del reloj (m). Por tanto, esta desviación (o skew) de (m) del reloj en un nodo A respecto a un reloj en el nodo B en el tiempo $t_0 + t$, puede calcularse como:

$$m_A^B(t_0 + t) = m_A^B(t_0) + \int_0^t \eta(u) \cdot du \qquad (1)$$

donde $\eta(u) \sim N(0, \sigma_{RW}^2)$, siendo σ_{RW} la desviación estándar característica del RW. Para estos relojes de bajo

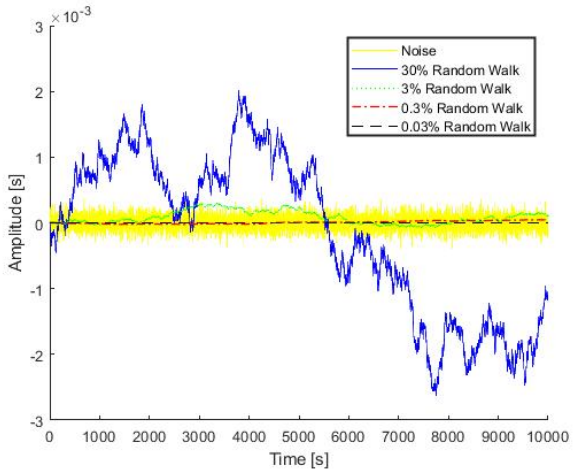


Figura 2: Representación de diferentes señales con ruido de fondo (amarillo) con deriva de 100 ppm, añadiendo efecto "random walk" con distintos % desde 30 hasta 0.03 de σ_{RW} vs ruido con σ_C.

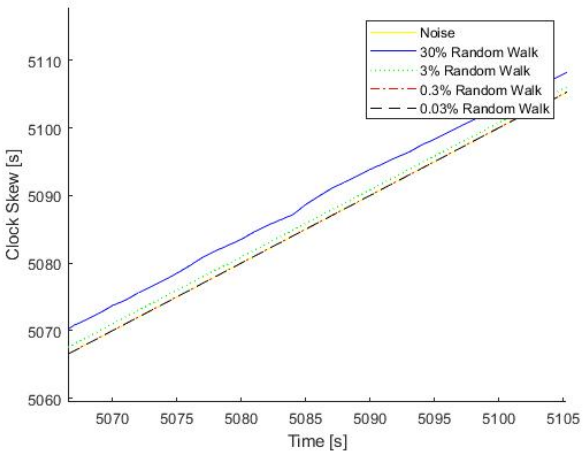


Figura 3: Efecto del "random walks" en la pendiente de un modelo linea de los relojes en segundos.

coste, se asume que σ_{RW} es del orden de 10^{-8} y 10^{-9} [8] con 1 ppm a 100 ppm [9].

Usando el siguiente código Matlab, podemos modelar un RW para $y(t) = m \cdot x(t)$ e incluyendo el efecto del ruido como:

```
%inicialización
 rw= zeros(1,longitud);
 ruido_w=normrnd(0,sigma_rw,1,longitud);
 ruido=normrnd(0,sigma_c,1,longitud);

%construyendo el RW
 for i=2:longitud
   rw(1,i)=rw(1,i-1)+ruido_w(1,i);
 end

%modelo de tiempo añadiendo RW y ruido
 y= x + rw.*x + ruido;
```

donde `normrnd` genera una matriz $1 \times longitud$ con distribución $N(0, \sigma)$ y `rw` denota la variable para generar el paseo aleatorio aplicado a la pendiente de `y`. Además, el ruido añadido asociado a los retrasos aleatorios de las comunicaciones (`ruido`), se modela como una distribución $N(0, \sigma_C^2)$, siendo σ_C la desviación estándar asociada a la suma de estos retrasos.

En la Fig. 2 se muestra el efecto de RW y la desviación que se crea con diferentes contribuciones, % de σ_C, desde el 30 % hasta el 0,03 %, respecto al ruido. En la Fig. 3 se muestra este efecto en el tiempo sobre la pendiente del reloj entre dos nodos.

IV. PROTOCOLO SIMPLE DE SINCRONIZACIÓN *ONE WAY*

La sincronización temporal se consigue mediante el intercambio de marcas de tiempo (time stamps) de los relojes locales, enviadas dentro de paquetes entre los nodos. Hay varios enfoques para estos intercambios, pero en este caso utilizaremos un enfoque simple, un enfoque unidireccional, en el que un nodo, llamado maestro, enviará paquetes de difusión con sus marcas de tiempo locales al resto de nodos (esclavos). Entonces, los esclavos pueden ajustar sus tiempos locales $(y(t_1), y(t_2), ..y(t_i))$ a los proporcionados por el maestro $(x(t_1), x(t_2), ...x(t_i))$, siendo i el número del paquete. A partir de estas marcas de tiempo, podemos aplicar diferentes técnicas de regresión para mejorar la sincronización temporal reduciendo el número de paquetes enviados y así predecir las siguientes marcas de tiempo, denotadas como $\hat{y}_{i+1}$. En este caso, el error de predicción viene dado por $|e_{i+1}| = |y_{i+1} - \hat{y}_{i+1}|$.

Nuestro objetivo es proporcionar un mecanismo para la estimación del tiempo, teniendo en cuenta las marcas de tiempo anteriores, así como la frecuencia utilizada para intercambiar estos paquetes. Esta frecuencia se conoce como periodos de sincronización (SP). Con este propósito, consideraremos diferentes parámetros para el análisis del error de sincronización: SP, tamaños de ventana utilizados para las técnicas de regresión, así como el grado de la misma. Respecto al SP, por un lado, cuanto más rápido intercambiemos paquetes, mayor será la precisión porque el reloj del esclavo no se desviará en exceso respecto al del maestro. Sin embargo, tiene un impacto en el consumo de energía porque estamos enviando más paquetes. Por otro lado, si queremos ahorrar energía reduciendo el número de paquetes intercambiados, la precisión del tiempo se ve afectada negativamente.

Además, el tamaño de la ventana utilizada para la estimación del tiempo, también influye en la minimización del error de sincronización, debido a los procesos aleatorios subyacentes en las desviaciones del reloj en estos nodos, así como el ruido introducido por los procesos de comunicación. Por tanto, dado que existen dos mecanismos subyacentes diferentes e independientes que influyen en estas regresiones, veremos diferentes comportamientos en función de σ_{RW} y σ_C. Por un lado si σ_C es mucho mayor que σ_{RW}, entonces se prefieren tamaños de ventana más

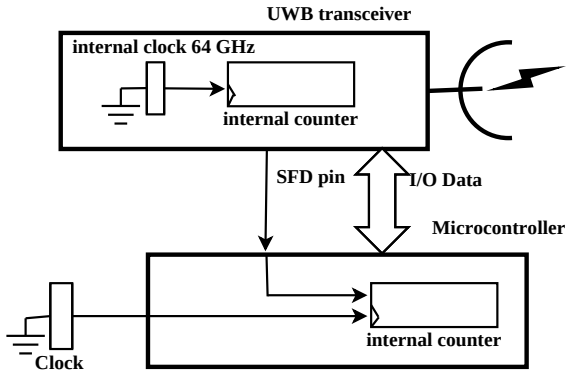


Figura 4: Detalle del transceptor DWM1000 UWB conectado al Arduino MKRZero.

grandes para la regresión. Por otro lado, si σ_{RW} es uno o dos órdenes de magnitud menor que σ_C, entonces se prefieren tamaños de ventana más cortos. Si ambos son similares o si σ_{RW} es mayor que σ_C, entonces la última muestra será la mejor estimación. En este contexto, existe un tamaño de ventana óptimo que minimizará el error para la estimación del tiempo [18].

A partir del comportamiento sintetizado del RW descrito en la Sección III, hemos evaluado experimentalmente el tamaño de ventana óptimo en función de la contribución del RW frente al ruido subyacente, midiendo ambos, en relación a σ_{RW} y σ_C, respectivamente. En la Tabla I, se muestran los tamaños de ventana óptimos para tres escenarios en los que se emplea una regresión lineal: a) σ_{RW} es mucho menor que σ_C, b) σ_{RW} es menor que σ_C, y c) σ_{RW} es similar o mayor que σ_C. Estos resultados se han determinando fijando un intervalo de confianza del 95 % para su significación estadística.

Obsérvese que cuando la contribución predominante es el ruido, entonces, como era de esperar, el tamaño óptimo de la ventana es lo más grande posible. Por contra, cuando predomina el RW, entonces el tamaño óptimo de la ventana es en la práctica la última marca de tiempo.

Tabla I: Tamaños de ventana óptimos para regresión lineal considerando RW simultáneo con ruido, con desviaciones σ_{RW} y σ_C respectivamente.

Escenario	% de RW	Ventana óptima
$\sigma_{RW} << \sigma_C$	$< 0{,}3$	$16 - \infty$
$\sigma_{RW} < \sigma_C$	$0{,}3 - 30$	3-15
$\sigma_{RW} >= \sigma_C$	> 30	2-3

V. PROTOTIPO DE NODO INALÁMBRICO

En la Fig. 4 se presenta un esquemático del nodo inalámbrico empleado en este trabajo, basado en el transceptor UWB Decawave DWM1000 [2]. El microcontrolador (o MCU) utilizado es Arduino MKRZero [19] de 32 bit basado en ARM@48 MHz con 32 KB de RAM y 256 KB de Flash.

Cabe destacar que configurando adecuadamente el MCU es posible conseguir un reloj con una frecuencia de hasta 96 MHz, y que se dispone de una librería específica para configurar este transceptor [20].

En la Fig. 4, se muestra como tanto el transceptor como la MCU tienen su propio reloj local, utilizando osciladores de cristal de cuarzo de bajo coste independientes. Hay que tener en cuenta que con este transceptor, utilizando un reloj externo de 38,4 MHz y la línea de sincronización (SYNC), podemos conseguir mejorar la sincronización entre ambos, como se muestra en [15]. Sin embargo, dificultaría el diseño ya que nuestra intención es definir un nodo inalámbrico sencillo.

Este transceptor UWB tiene un contador interno de 40 bits conectado a un reloj interno de 64 GHz. Este contador se utiliza para marcar el tiempo de la trama, tanto al transmitir como al recibir. Utilizando el SFD, el transceptor emisor puede sincronizar la marca de tiempo cuando se envía el campo SFD de la trama, y el transceptor receptor puede sincronizar su marca de tiempo, justo cuando se detecta el SFD. Además, para la sincronización externa, el transceptor proporciona un pin SFD, como se muestra en la Fig. 4, activado según el mecanismo SFD, pero sólo en la recepción de tramas (no en la transmisión), permitiendo en la MCU el marcado de tiempo de las tramas cuando se detecta este campo. Por tanto en la práctica, la activación del pin SFD limitará la sincronización con la MCU. Obsérvese que como este transceptor no fue diseñado para la sincronización, su caracterización no se encuentra en los manuales y/o especificaciones

VI. RESULTADOS

En esta sección, para analizar el error de sincronización en las estimaciones de tiempo, utilizaremos el valor medio de los errores absolutos de predicción, o Mean Absolute Prediction Error (*MAPE*) como se describe en la Sección IV. Para los experimentos utilizamos el protocolo descrito en la Sección IV y colocamos el maestro a un metro de distancia de los esclavos en una habitación con una temperatura estable en torno a los 25ºC. El maestro envía paquetes de sincronización con las marcas de tiempo correspondientes a los envíos de su transceptor UWB, y los esclavos, utilizando el mecanismo SFD, determinarán el tiempo de estos paquetes tanto en el transceptor UWB como en la MCU. Durante las pruebas, el maestro envió 94080 paquetes con un SP de 200 ms.

Además, para evaluar el efecto del RW en los desvíos del reloj, hemos testeado el protocolo de sincronización utilizando diferentes SP, tal y como se comenta en la sección IV. Hemos utilizado varios SP: 200 ms, 1 s, 10 s, 30 s y 60 s. En las Fig. 5-9 se muestran los *MAPE* calculados a partir de las marcas de tiempo registradas en el UWB. Además, en estas figuras hemos evaluado diferentes técnicas de regresión, utilizando aproximaciones lineales, cuadráticas, cúbicas y cuárticas, así como modificando el tamaño de la ventana aplicada en la regresión.

Como era de esperar, el menor *MAPE* viene dado por la utilización del SP más corto, con 200 ms, y utilizando la regresión lineal, como se muestra en la Fig. 5. En el eje x se da el índice del tamaño de la ventana para cada

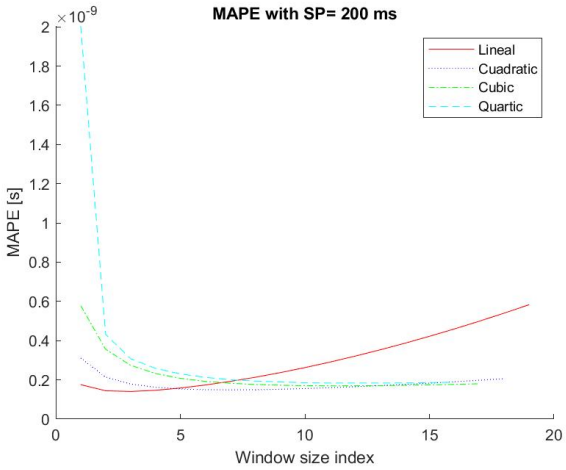


Figura 5: MAPE para un SP de 200 ms utilizando diferentes regresiones.

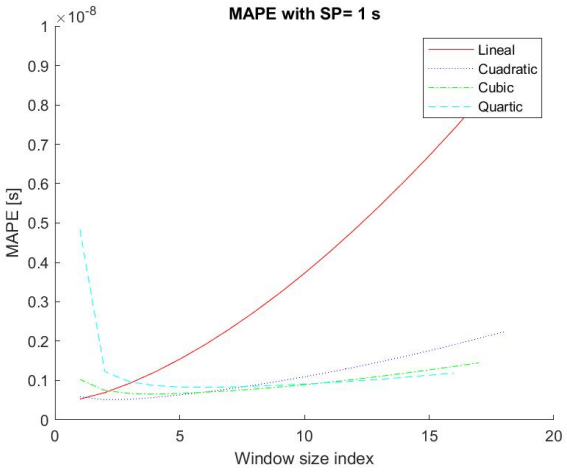


Figura 6: MAPE para un SP de 1 s utilizando diferentes regresiones.

regresión. Para una regresión lineal, el primer índice comienza con un tamaño de ventana mínimo de 2 muestras. Para las aproximaciones cuadrática, cúbica y cuártica los tamaños mínimos de ventana son de 3, 4 y 5 muestras respectivamente, el grado más uno. Entonces, según Fig. 5, el óptimo se consigue con un índice de tamaño de ventana de 3, es decir un tamaño de ventana de 4 muestras, con un *MAPE* de 149 ps.

Con un SP de 1 s, el óptimo se consigue con un tamaño de ventana en el tercer índice utilizando la regresión cuadrática, es decir, un tamaño de ventana de 4 muestras, ya que el índice de ventana óptimo es 2 y para esta regresión, el tamaño de ventana mínimo es 3, como se muestra en las Fig. 6, con un *MAPE* de 511,8 ps. Para SP más altos (10, 30 y 60 s), como se muestra en la Fig. 7-9, siempre se consiguen los mejores resultados con una regresión cuadrática y con el menor tamaño de ventana posible. Esto se debe a la prevalencia del efecto RW, como se ha comentado anteriormente. En la Tabla II, resumimos los valores óptimos utilizando las marcas de tiempo del UWB, indicando el SP, *MAPE*, grado (D) y tamaño de ventana (WS).

Tabla II: Resumen de los valores óptimos, con el valor mínimo de *MAPE* [ns], según SP, grado (D) y tamaño de ventana (WS) con las marcas temporales de UWB.

SP [s]	0.2	1	10	30	60
MAPE [ns]	0,141	0,512	16,56	95,5	283,4
D	1	2	2	2	2
WS	4	4	3	3	3

En las Fig. 10-11 se muestran los histogramas del error cuando se utiliza un SP de 200 ms tanto con regresión lineal como cuadrática respectivamente. En este caso, utilizando la regresión lineal la precisión aumenta en comparación con las otras técnicas de regresión.

Por último, en las Fig. 12-16 se muestran los errores basados en las marcas de tiempo de la MCU, utilizando los mismos escenarios y configuraciones. A primera vista,

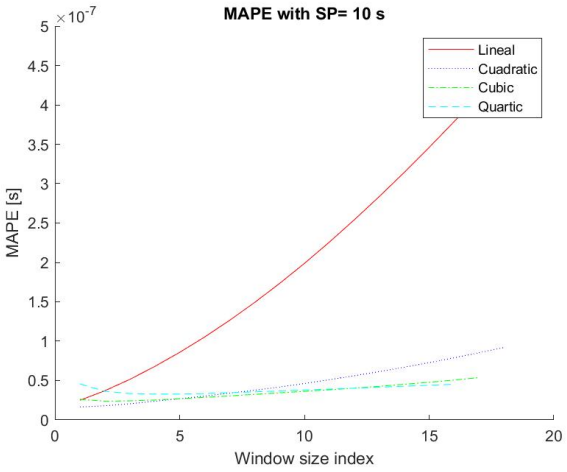


Figura 7: MAPE para un SP de 10 s utilizando diferentes regresiones.

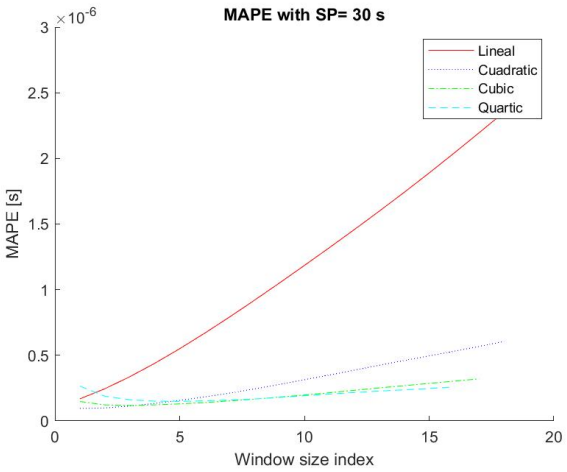


Figura 8: MAPE para un SP de 30 s utilizando diferentes regresiones.

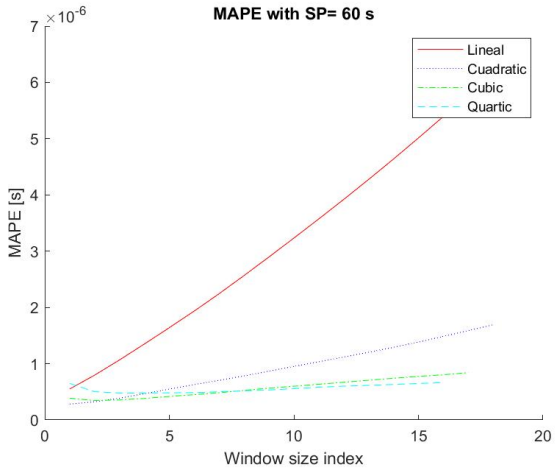


Figura 9: MAPE para un SP de 60 s utilizando diferentes regresiones.

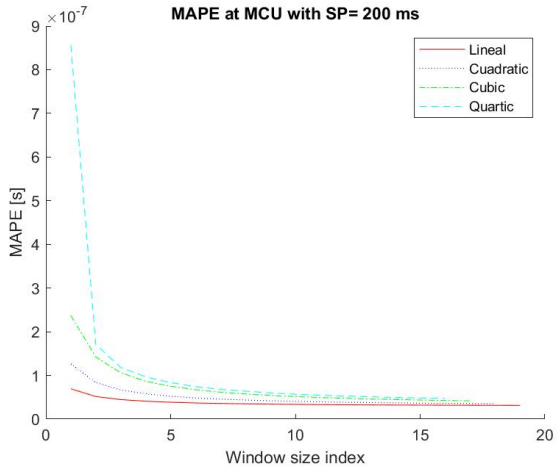


Figura 12: MAPE para un SP de 200 ms utilizando diferentes regresiones en la MCU.

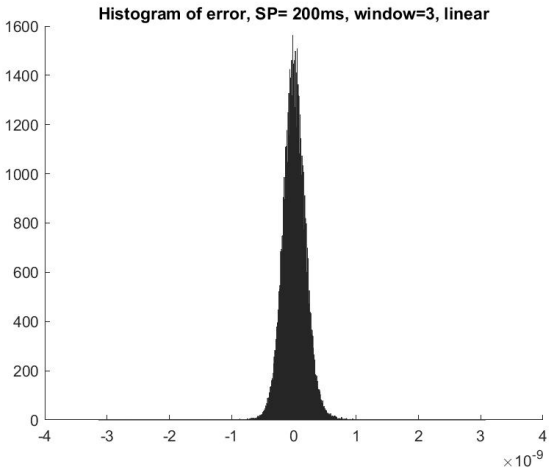


Figura 10: Histograma del error utilizando regresión lineal con un SP de 1s y una ventana de 3.

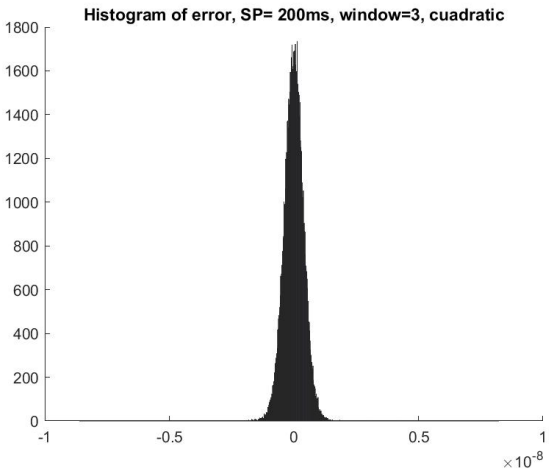


Figura 11: Histograma del error utilizando regresión cuadrática con un SP de 1s y una ventana de 3.

observamos que los errores aumentan en comparación con los anteriores, obtenidos utilizando la marca de tiempo interna en el transceptor UWB. Como hemos comentado, esto se debe a que el marcado de tiempo SFD en la MCU se basa en la activación de la salida del pin SFD. Además, los contadores internos de la MCU funcionan a una velocidad inferior a la del transceptor UWB. Mientras que el transceptor UWB trabaja a 64 GHz, el reloj de la MCU trabaja a 78,698 MHz y no 96 MHz, como se explica en la Sección V, porque si utilizamos una frecuencia más alta, el MCU empieza a perder paquetes.

Cabe mencionar, que como la MCU está utilizando una frecuencia de reloj más baja en comparación con UWB, el comportamiento de RW es diferente, como podemos ver en las Fig. 12-13. En este caso, el tamaño de ventana óptimo viene dado por la mayor ventana posible del rango utilizado (20), tendiendo a infinito, con un *MAPE* de 31,35 con un SP de 0,2 s y empleando una regresión lineal y de 38,51 ns con un SP 1 s y empleando una regresión cuadrática. Por lo tanto, el efecto RW es menor en comparación con el ruido. Sin embargo, al aumentar el SP a 10, 30 y 60 s, el efecto RW aparece y el tamaño de ventana óptimo se encuentra por debajo de 5. Además, para los diferentes SP, las regresiones óptimas vienen dadas por la cuadrática, excepto con un SP de 0,2 s donde la regresión óptima es lineal. En la Tabla III, resumimos para las marcas de tiempo en la MCU los valores óptimos, indicando el SP, *MAPE*, D y WS utilizados.

Tabla III: Resumen de los valores óptimos, con el valor mínimo de *MAPE* [ns], según SP, grado (D) y tamaño de ventana (WS) con las marcas temporales de MCU.

SP [s]	0,2	1	10	30	60
MAPE [ns]	31,35	38,51	101,6	337,5	761,1
D	1	2	2	2	2
WS	∞	∞	2	4	4

En la práctica, hay que tener en cuenta que si no se conociera la posición exacta de los nodos, podríamos cal-

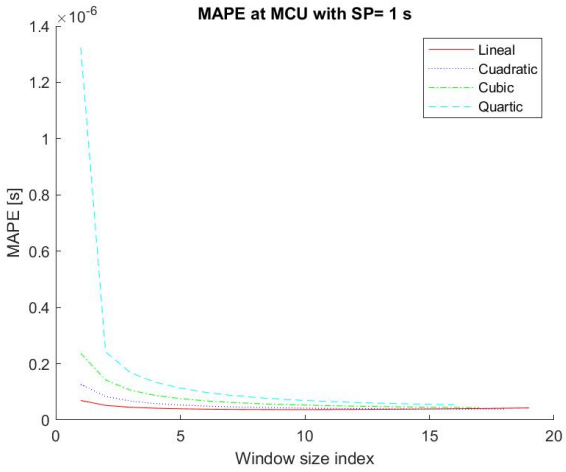


Figura 13: MAPE para un SP de 1 s utilizando diferentes regresiones en la MCU.

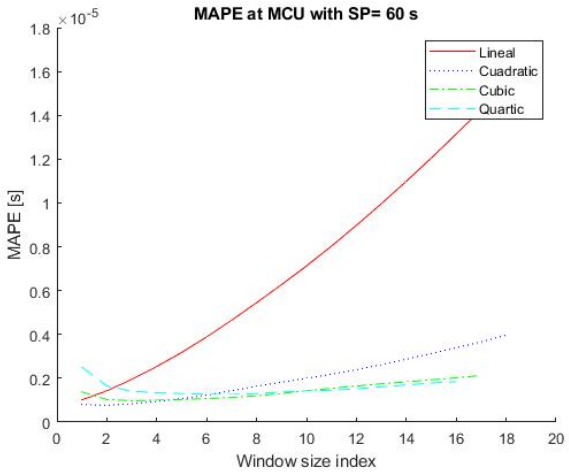


Figura 16: MAPE para un SP de 60 s utilizando diferentes regresiones en la MCU.

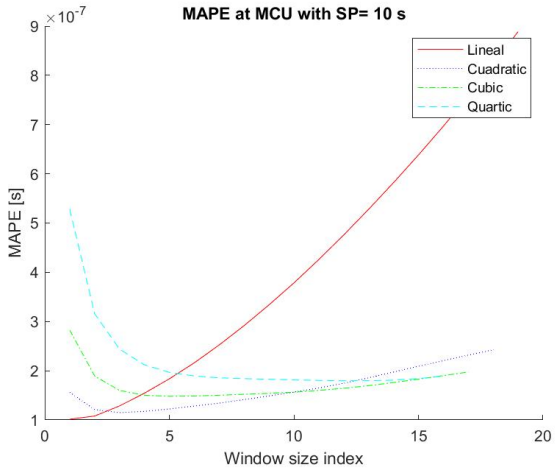


Figura 14: MAPE para un SP de 10 s utilizando diferentes regresiones en la MCU.

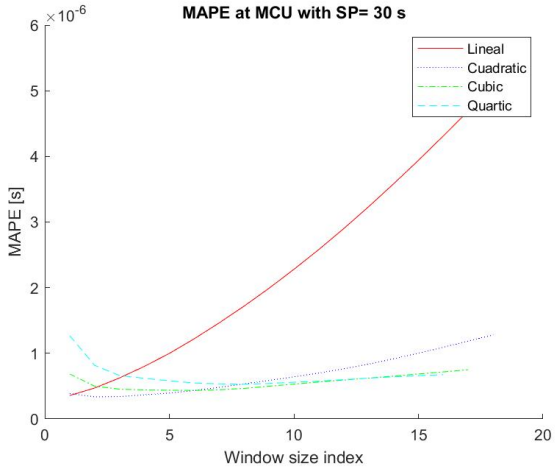


Figura 15: MAPE para un SP de 30 s utilizando diferentes regresiones en la MCU.

cularla fácilmente con el mismo transceptor, midiendo el TOF entre ellos. El TOF afectará a la precisión del tiempo como una compensación constante. En el transceptor, el TOF se puede calcular fácilmente utilizando la expresión propuesta en [21] por el equipo Decawave, basado en el método *Poll-Response-Final* [22]. Este transceptor para aplicaciones de alcance tiene una inexactitud del orden de cm [2], que a la velocidad de la luz supone un desplazamiento constante de 33 ps por cm.

VII. CONCLUSIONES Y TRABAJO FUTURO

Las comunicaciones UWB pueden integrarse dentro de los nodos tradicionales de las WSN, manteniendo la característica de bajo coste de sus nodos. En este trabajo, utilizando este tipo de comunicaciones, hemos demostrado que podemos conseguir una sincronización en una escala de tiempo por debajo de los nanosegundos utilizando componentes comerciales disponibles, en particular utilizando un nodo inalámbrico sencillo basado en un transceptor UWB Decawave DWM1000 conectado a un MCU Arduino MKRZero. Cada uno de ellos con un oscilador y un sistema de reloj propios.

Hemos considerado el comportamiento de la deriva (o skew) de los relojes modelado como un proceso RW tanto entre los transceptores UWB, como entre el transceptor y la MCU. En base a ello y utilizando experimentos reales, hemos analizado el tamaño de ventana óptimo que minimiza el error de estimación para la sincronización temporal utilizando diferentes regresiones y el protocolo de sincronización *one way*. En el transceptor UWB, dado que utilizamos un reloj de 64 GHz, las derivas del RW tienen más influencia que el ruido intrínseco añadido en los intercambios de paquetes y sus marcas de tiempo comparado con la MCU. Uno de los principales inconvenientes del transceptor UWB DWM1000 es que no proporciona un pin de salida SFD cuando transmite tramas, sólo cuando recibe tramas. Debemos tener en cuenta que este transceptor fue diseñado principalmente para aplicaciones de localización y no para proporcionar una sincronización precisa entre

relojes, que es el propósito de la investigación presentada en este trabajo.

En cuanto a los resultados, cuando el marcado de tiempo SFD se realiza en el propio transceptor tanto en la transmisión como en la recepción, logramos un error de sincronización mínimo (*MAPE*) de 149 ps y 511,8 ps con un SP de 0,2 y 1 s, respectivamente. Sin embargo, cuando utilizamos las marcas de tiempo en la MCU a través del pin SFD, logramos un *MAPE* de 31,35 ns y 38,51 ns considerando SP de 0,2 s y 1 s, respectivamente.

Esto confirma que podemos explotar como trabajo futuro las características del transceptor DWM1000 para mejorar la sincronización temporal en WSN, utilizando protocolos de sincronización más complejos con el fin de conseguir una mayor precisión, así como hacer uso de esta sincronización en aplicaciones tales como localización de eventos sonoros y beamforming colaborativo.

AGRADECIMIENTOS

Este trabajo ha sido posible a la financiación de la Conselleria de Innovacion, Universidades, Ciencia y Sociedad Digital de la Generalitat Valenciana através de AEST/2021/16, BEST/2021/150, AICO/2020/154, la Universitat de València con UV-INV-AE-1544281. Y BIA2016-76957-C3-1-R y BES-2017-082340 financiadas por MCIN/AEI/ 10.13039/501100011033 y por "FEDER Una manera de hacer Europa" y "FSE Invierte en tu futuro".

REFERENCIAS

[1] E. Navarro-Camba, S. Felici-Castell, J. Segura-García, M. García-Pineda, and J. Pérez-Solano, "Feasibility of a stochastic collaborative beamforming for long range communications in wireless sensor networks," *Electronics*, vol. 7, no. 12, p. 417, Dec 2018. [Online]. Available: http://dx.doi.org/10.3390/electronics7120417

[2] Decawave, "DWM1001 System Overview And Performance," https://www.Decawave.com/content/dwm1001-system-overview-andperformance, 2015, accessed: 26/05/2021.

[3] J. J. Pérez-Solano, S. Ezpeleta, and J. M. Claver, "Indoor localization using time difference of arrival with uwb signals and unsynchronized devices," *Ad Hoc Networks*, vol. 99, p. 102067, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S157087051930215X

[4] S. Ganeriwal, I. Tsigkogiannis, H. Shim, V. Tsiatsis, M. Srivastava, and D. Ganesan, "Estimating Clock Uncertainty for Efficient Duty-Cycling in Sensor Networks," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 3, pp. 843–856, 2009. [Online]. Available: http://dx.doi.org/10.1109/TNET.2008.2001953

[5] M. Maróti, B. Kusy, G. Simon, and A. Lédeczi, "The flooding time synchronization protocol," in *Proceedings of the 2nd international conference on Embedded networked sensor systems*, ser. SenSys '04. New York, NY, USA: ACM, 2004, pp. 39–49. [Online]. Available: http://doi.acm.org/10.1145/1031495.1031501

[6] S. Yoon, C. Veerarittiphan, and M. L. Sichitiu, "Tiny-sync: Tight time synchronization for wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 3, no. 2, Jun. 2007. [Online]. Available: http://doi.acm.org/10.1145/1240226.1240228

[7] T. Schmid, R. Shea, Z. Charbiwala, J. Friedman, M. B. Srivastava, and Y. H. Cho, "On the interaction of clocks, power, and synchronization in duty-cycled embedded sensor nodes," *ACM Trans. Sen. Netw.*, vol. 7, no. 3, pp. 24:1–24:19, Oct. 2010. [Online]. Available: http://doi.acm.org/10.1145/1807048.1807053

[8] Z. Zhong, P. Chen, and T. He, "On-demand time synchronization with predictable accuracy," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 2480–2488. [Online]. Available: http://dx.doi.org/10.1109/INFCOM.2011.5935071

[9] Seiko Instruments Inc, "Product List," http://speed.sii.co.jp/pub/compo/quartz/productListEN.jsp, 2014, accessed: 06/02/2021.

[10] J. J. Pérez-Solano and S. Felici-Castell, "Adaptive time window linear regression algorithm for accurate time synchronization in wireless sensor networks," *Ad Hoc Networks*, vol. 24, pp. 92–108, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.adhoc.2014.08.002

[11] F. Spitzer, *Principles of Random Walk*, ser. Graduate Text in Mathematics. Springer, 2001.

[12] S. Ganeriwal, R. Kumar, and M. B. Srivastava, "Timing-sync protocol for sensor networks," in *Proceedings of the 1st international conference on Embedded networked sensor systems*, ser. SenSys '03. New York, NY, USA: ACM, 2003, pp. 138–149. [Online]. Available: http://doi.acm.org/10.1145/958491.958508

[13] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, april 2005, pp. 364 – 369. [Online]. Available: http://dx.doi.org/10.1109/IPSN.2005.1440950

[14] F. Bonafini, P. Ferrari, A. Flammini, S. Rinaldi, and E. Sisinni, "Exploiting time synchronization as side effect in uwb real-time localization devices," in *2018 IEEE International Symposium on Precision Clock Synchronization for Measurement, Control, and Communication (ISPCS)*, 2018, pp. 1–6.

[15] A. Dongare, P. Lazik, N. Rajagopal, and A. Rowe, "Pulsar: A wireless propagation-aware clock synchronization platform," in *2017 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 2017, pp. 283–292.

[16] S. Niranjayan and A. F. Molisch, "Ultra-wide bandwidth timing networks," in *2012 IEEE International Conference on Ultra-Wideband*, 2012, pp. 51–56.

[17] Y. Xie, G. J. M. Janssen, and A. van der Veen, "A practical clock synchronization algorithm for uwb positioning systems," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3891–3895.

[18] J. J. Pérez-Solano and S. Felici-Castell, "Improving time synchronization in wireless sensor networks using bayesian inference," *Journal of Network and Computer Applications*, vol. 82, pp. 47 – 55, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804517300164

[19] Arduino, "MKR ZERO," https://store.arduino.cc/arduino-mkr-zero-i2s-bus-sd-for-sound-music-digital-audio-data, 2018, accessed: 26/07/2021.

[20] Thomas Trojer, "Github: Library that offers basic functionality to use Decawave's DW1000 chips/modules with Arduino," https://github.com/thotro/arduino-dw1000, 2018, accessed: 28/11/2021.

[21] D. Neirynck, E. Luk, and M. McLaughlin, "An alternative double-sided two-way ranging method," in *2016 13th Workshop on Positioning, Navigation and Communications (WPNC)*, 2016, pp. 1–4.

[22] Decawave, "The implementation of two-way ranging with the DW1000," https://www.decawave.com/wp-content/uploads/2018/10/APS013_The-Implementation-of-Two-Way-Ranging-with-the-DW1000_v2.3.pdf, 2018, accessed: 26/11/2021.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

URBAURAMON: Herramientas inteligentes para la gestión y monitorización acústica

J. Lopez-Ballester*, S. Felici-Castell*, J. Segura-Garcia*, JJ. Perez-Solano*, A. Soriano-Asensi*,
* ETSE, Universitat de València
jesus.lopez-ballester@uv.es, felici@uv.es, jsegura@uv.es, jjperezs@uv.es, soan@uv.es

***Palabras Clave*—redes dedicadas, sound scape, wireless sensor networks, neural networks, acoustic parameters**

En el proyecto URBAURAMON se han desarrollado herramientas que permiten la evaluación y análisis subjetivo de la exposición de la población al sonido ambiental, así como la medición de parámetros acústicos para determinar la impresión perceptual de los entornos.

El paisaje sonoro es un elemento fundamental para el confort de las personas y regulado por *Environmental Noise Directive* (END) 2002/49/EC e ISO 12913, basado en parámetros como *Loudness* (L), *Sharpness* (S), *Fluctuation Strength* (F) y *Roughness* (R) que determinan la molestia subjetiva o *Psychoacoustic Annoyance*. En este contexto, las redes de sensores inalámbricas acústicas (*Wireless Acoustic Sensor Networks* (WASN)) son una opción muy interesante [1], pero el coste computacional que conlleva el procesamiento y monitorización del paisaje sonoro, así como la medición de parámetros acústicos para determinar la impresión perceptual de los entornos (tiempos de reverberación y claridad (ISO 3382), índices de inteligibilidad *(Speech Transmission Index)/(Speech Intelligibility Index)* (STI/SII) definidos en EC 60268-16, ISO 9921 y ANSI S3.5-1997), dificultan su implementación en nodos basados en *Single Board Computers* (SBC).

El proyecto URBAURAMON (http://www.uv.es/urbauramon/) ha tenido como objetivo desarrollar herramientas para ayudar en este proceso de monitorización acústica. En particular, la técnicas utilizadas para implementar estos parámetros se han basado en técnicas de virtualización utilizando contenedores y su orquestación [2], junto con técnicas basadas en redes neuronales (*deep Convolutional Neural Networks* (CNNs)) para reducir el coste computacional para el cálculo de estos parámetros, con un error relativo inferior al 3 % y una velocidad 200 veces superior a la del cálculo directo mediante procesado de señales convencional [3][4].

Además, las tecnologías 5G IoT han permitido flexibilizar esta arquitectura, ya que permiten descargar computacionalmente (*computational offloading*) la carga en los nodos, que las tecnologías anteriores a 5G no ofrecían. En particular, se han analizado diferentes técnicas de descarga mediante el desarrollo de diferentes *functional splittings*, divisiones funcionales, de los algoritmos de los parámetros acústicos [5] utilizando técnicas de *network slicing*. Dicha arquitectura está basada en 5G *Mobile/Multi-access Edge Computing* (MEC).

AGRADECIMIENTOS

Este proyecto ha sido realizado bajo la financiación de la Generalitat Valenciana con el proyecto GV/2020/052, AICO/2020/154, AEST/2020/048, BEST/2021/150, AEST/2021/016 y la Universitat de València con la Acción Especial UV-INV-AE-1544281 y BIA2016-76957-C3-1-R y BES-2017-082340 financiadas por MCIN/AEI/ 10.13039/501100011033 y por “FEDER Una manera de hacer Europa” y “FSE Invierte en tu futuro”.

REFERENCIAS

[1] A. Pastor-Aparicio, J. Segura-Garcia, J. Lopez-Ballester, S. Felici-Castell, M. García-Pineda, and J. J. Pérez-Solano, “Psychoacoustic annoyance implementation with wireless acoustic sensor networks for monitoring in smart cities,” *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 128–136, 2020.

[2] R. Fayos-Jordan, S. Felici-Castell, J. Segura-Garcia, J. Lopez-Ballester, and M. Cobos, “Performance comparison of container orchestration platforms with low cost devices in the fog, assisting internet of things applications,” *Journal of Network and Computer Applications*, vol. 169, p. 102788, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804520302605

[3] J. Lopez-Ballester, A. Pastor-Aparicio, S. Felici-Castell, J. Segura-Garcia, and M. Cobos, “Enabling real-time computation of psychoacoustic parameters in acoustic sensors using convolutional neural networks,” *IEEE Sensors Journal*, vol. 20, no. 19, pp. 11 429–11 438, 2020.

[4] J. Lopez-Ballester, J. M. Alcaraz Calero, J. Segura-Garcia, S. Felici-Castell, M. Garcia-Pineda, and M. Cobos, “Speech intelligibility analysis and approximation to room parameters through the internet of things,” *Applied Sciences*, vol. 11, no. 4, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/4/1430

[5] J. Segura-Garcia, J. M. A. Calero, A. Pastor-Aparicio, R. Marco-Alaez, S. Felici-Castell, and Q. Wang, “5g iot system for real-time psycho-acoustic soundscape monitoring in smart cities with dynamic computational offloading to the edge,” *IEEE Internet of Things Journal*, pp. 1–1, 2021.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Sistema IoT para la monitorización de parámetros de sala e inteligibilidad del habla

J. Lopez-Ballester*, S. Felici-Castell*, J. Segura-Garcia*, JJ. Perez-Solano*, A. Soriano-Asensi*,
* ETSE, Universitat de València
jesus.lopez-ballester@uv.es, felici@uv.es, jsegura@uv.es, jjperezs@uv.es, soan@uv.es

Resumen

En los últimos años las Redes Inalámbricas de Sensores Acústicos (WASN) se han utilizado de manera extendida en diferentes campos acústicos tanto en ambientes interiores como exteriores. Muchas de estas aplicaciones se orientan a la localización o identificación de fuentes sonoras y medir parámetros acústicos específicos del entorno en cuestión. En este artículo, estudiamos el diseño y la aplicación de un sistema IoT formado por una WASN que permite monitorizar los diferentes parámetros acústicos de una sala que nos darán una idea de las características acústicas de la misma, pero de una forma más rápida y económica que los métodos tradicionales. Como base para los nodos de la WASN se ha utilizado un conjunto de Raspberry Pi 3 (RPi), una para el control, procesado y reproducción de diferentes señales acústicas y cuatro para grabar en diferentes puntos de la sala simultáneamente. Las señales grabadas se emplean para calcular tanto la respuesta impulsiva (Impulsive Response, IR), como diferentes parámetros acústicos de sala, además de parámetros de inteligibilidad como el Speech Intelligibility Index (SII) y se almacenarán en una base de datos local en caso de que no haya conexión a la nube. Por último se han realizado diversas mediciones en espacios reales para poner a prueba la viabilidad y funcionamiento del sistema. De este modo, se han explorado tanto la evaluación de los parámetros acústicos de la sala a partir de mediciones asíncronas de IR como la monitorización contínua mediante un sistema IoT en un entorno real.

***Palabras Clave*—IoT, WASN, Acústica de salas, respuesta impulsiva, índice de inteligibilidad del habla, estimación de parámetros de sala.**

I. INTRODUCCIÓN

Comenzando con la publicación de Sabine [1] en el campo del estudio de la reverberación acústica en las salas, diversos investigadores han enfocado su carrera a la medición del comportamiento acústico de los entornos, sobretodo interiores. Esto llevó a la estandarización de los procesos de medida de ciertos parámetros acústicos concretos en la ISO 3382 [2, 3, 4]. Estos procedimientos que han sufrido revisiones y actualizaciones como es lógico, pero manteniendo siempre algunos aspectos a la libre elección del usuario, para permitir la innovación y la investigación al respecto.

Así pues se definen una serie de parámetros acústicos que permiten la descripción de una sala como pueden ser por ejemplo el Tiempo de Reverberación con caída de 60 dB (RT60), la definición (D), la claridad de la voz (C50) o la sonoridad (G), recogidos en la norma ISO-3382. Si estos parámetros están relacionados con la reverberación y con la distribución energética del sonido, en la norma ISO-9921 [5] y ANSI S3.5-1997 [6] se definen otros relacionados específicamente con la transmisión e inteligibilidad del habla, como son el Speech Transmission Index (STI) o el Speech Intelligibility Index (SII).

Como la mayoría de estos parámetros se calculan a partir de la respuesta impulsiva de la habitación, el procedimiento habitual consiste en emitir una señal de audio por un altavoz al mismo tiempo que se graba en diferentes posiciones que nos interesen de una sala para extraer la respuesta impulsiva de la sala. Este proceso se complica en salas de dimensiones considerables, puesto que la distancia entre la fuente emisora de sonido y los micrófonos puede ser muy grande, lo que dificulta la instalación, alimentación y cableado del sistema de medición. Las Redes Inalámbricas de Sensores Acústicos (WASN) pueden resolver este problema al simplificar el proceso de instalación y funcionamiento gracias a sensores que se alimentan de forma autónoma y se comunican de forma inalámbrica.

Las WASN han permitido la automatización técnica del proceso de toma de medidas, según indicaciones específicas acordes a las normativas pertinentes. No obstante, la investigación en este campo es fundamental para mejorar los procedimientos de medición empleando por ejemplo diferentes técnicas de procesado de señal [7, 8, 9, 10 ,11]. También los protocolos de medida y comunicación empleados, pues usando WASNs se puede obtener la información acústica de una sala completa en una sola monitorización simultánea mediante sensores distribuidos en la misma. Finalmente, si dotamos a la WASN de conexión a internet se convierte en un sistema IoT de

monitorización acústica de salas, tal y como describimos en este artículo.

II. ESTADO DEL ARTE

En los últimos años, las WASNs se han empleado en muchos estudios relacionados con la acústica de espacios exteriores e interiores, centrados la mayor parte de las veces en la localización de fuentes [12, 13], el rastreo [14] o la identificación de las mismas [15]. También a la medición de características específicas del entorno estudiado [16]. En el estudio mostrado en [17], los autores utilizan una simulación basada en WASNs para estimar el rendimiento acústico de una sala, empleando para ello Respuestas Impulsivas de Sala (RIRs).

A los parámetros acústicos de sala mencionados anteriormente, RT60, D, C50, etc, podemos sumar el STI y el SII [18], como parámetros orientados específicamente a la transmisión oral de la palabra. Aunque estos parámetros nos hagan pensar en salas de concierto o auditorios, al estar orientados al habla en vez de a la música, es muy interesante también analizarlos en espacios docentes o dedicados a la atención médica, como un quirófano donde una buena transmisión e inteligibilidad de la palabra puede ser vital [19, 20, 21].

Así pues hemos seleccionado un conjunto discreto de 5 parámetros acústicos de sala como se puede ver en la Tabla I: RT60, C50, C80 , STI y SII. A partir de la respuesta impulsiva y lo que tarde en caer 60 dB, RT60 nos dará información del tiempo de reverberación de la sala. Los tiempos muy cortos son propios de salas de grabación anecóicas mientras que los tiempos largos son característicos de salas muy extensas y con materiales duros. C50 y C80 nos aportan información acerca de la distribución de la energía en la respuesta impulsiva, para saber si la energía se concentra antes de los 50 ms y antes de los 80 ms respectivamente, o si está muy dispersa a lo largo de la respuesta impulsiva. Valores superiores a 2 dB son deseables como mínimo en la mayoría de casos. STI nos da información de cómo se transmite el habla en una sala, y se obtiene de una manera resumida a partir de la suma ponderada de los Índices de Transferencia de la Modulación (MTI), que se obtienen aplicando diferentes modulaciones por bandas a la respuesta impulsiva, desde los 125Hz a los 8 kHz, teniendo en cuenta además factores relacionados con el sistema auditivo, la reverberación y el ruido. STI se presenta como un coeficiente con valores de 0 a 1 donde 0 representa una muy mala trasmisión del habla y 1 muy buena. SII nos aporta información de cómo es de inteligible el habla que se percibe en una posición de una sala. A diferencia de los demás parámetros, el SII se calcula a partir de una señal de habla grabada tal y cómo describe la norma ANSI S3.5-1997 [6]. De forma resumida se trata de la suma por bandas de la señal de Escucha de cada banda (A_i) ponderada por el índice de Importancia de cada banda (I_i), teniendo en cuenta también el nivel de ruido presente y el patrón estándar de audición empleado. Aunque el análisis se puede hacer en 6 bandas (1 por octava) hemos analizado 18 bandas, una por cada 1/3 de octava, de un conjunto de frecuencias del espectro audible total comprendido entre 160 y 8 kHz. SII será representado por un valor de 0 a 1 con la correspondencia mostrada en la Tabla II.

Estos parámetros permiten realizar un análisis acústico general a una sala, incluyendo parámetros de reverberación, energéticos y de inteligibilidad del habla. Todos ellos se calculan a partir de la respuesta impulsiva de la sala excepto el SII [6], que se puede obtener del audio de una persona hablando grabada en las ubicaciones del entorno en la que queremos estudiar la inteligibilidad. De esta forma se simplifica el procesado requerido para la señal y así el coste computacional de las plataformas hardware usadas. Teniendo en cuenta la época actual y los problemas sanitarios debidos al COVID19, la adaptación de los laboratorios y aulas no diseñadas en un principio para la docencia, hemos visto necesario incluir un sistema que permita caracterizar y medir de forma simple y rápida la inteligibilidad del habla en estos entornos, en diferentes puntos del recinto y poder actuar en consecuencia.

Por ello hemos centrado nuestro estudio en el SII principalmente, permitiendo además de en el cálculo de los demás parámetros citados sólo en caso de ser necesario. Como sucede con el cálculo de los parámetros de molestia psico-acústica como se vio anteriormente [22], el proceso de cálculo necesario para obtener los parámetros se sala es complejo, y por lo tanto largo de realizar en un nodo del sistema IoT, tomando una media de 1.7 segundos de tiempo por cada segundo de audio a analizar, de manera que por defecto los cálculos de RT60, C50, C80 y STI se realizan en el nodo de control, ya que requiere del cálculo previo de la respuesta impulsiva de la sala. Como el cálculo de SII no requiere de este cálculo previo, representa únicamente un 10 % de este tiempo. Por ello, por defecto nuestro sistema IoT está orientado a calcular SII en cada posición de los nodos de manera continua y de manera puntual puede medir los parámetros de sala empleando más tiempo, como es lógico en realizar los cálculos en el nodo de control. Por ello en un futuro trabajo nos proponemos diseñar y entrenar una red neuronal convolucional (CNN) que permita predecir los parámetros con un error de predicción bajo, menor del 3 %, tal y como hicimos también con los parámetros de molestia psico-acústica [22] para poder realizar los cálculos en el nodo de manera rápida y eficiente. Si bien el cálculo del SII se puede realizar de forma asíncrona, dado que cada nodo calculará un valor de SII según su ubicación, para el resto de parámetros se han grabado las señales de forma síncrona mediante la conexión WiFi de las RPi, sincronizando la emisión de los barridos en frecuencia emitidos con la grabación en los nodos mediante marcas de tiempo. Si bien esta sincronización no es demasiado precisa, fundamentalmente porque está basada en software y es susceptible de sufrir retrasos y variaciones debidos al acceso a la red, es más que suficiente para el propósito de este análisis, que nos permitirá tener una idea general del comportamiento acústico de la sala. Por lo tanto, en este trabajo exponemos el diseño y la implementación de

Cuadro I: Parámetros acústicos de sala, clasificación y requisitos

Parámetro	*Clasificación*	*Requerimiento*	*Normativa*
$RT60$ (Reverb. Time 60dB)	Reverberación	Respuesta Impulsiva	ISO 3382-2
$C50$ (Speech Clarity)	Energía Habla	Respuesta Impulsiva	ISO 3382-2
$C80$ (Music Clarity)	Energía Musical	Respuesta Impulsiva	ISO 3382-1
STI (Speech Transmission Index)	Inteligibilidad	Respuesta Impulsiva	ISO 9921
SII (Speech Intelligibility Index)	Inteligibilidad	Señal de audio	ANSI S3.5-1997

Cuadro II: Escala de valores de SII

SII	**Nivel de Inteligibilidad**
0.00 – 0.30	Muy mala
0.30 – 0.45	Mala
0.45 – 0.60	Aceptable
0.60 – 0.75	Buena
0.75 – 1.00	Muy buena

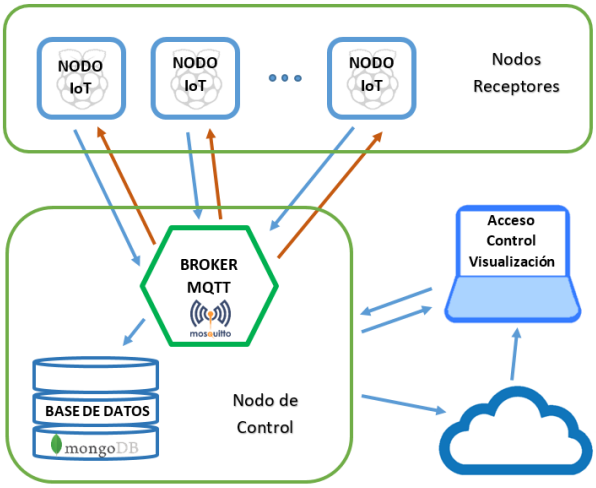


Figura 1: Esquema del sistema IoT.

un sistema IoT basado en RPi, para realizar mediciones acústicas síncronas y asíncronas orientadas a estudiar principalmente la inteligibilidad del habla en una sala o parámetro SII, además las características acústicas del espacio.

III. SISTEMA IOT DE MONITORIZACIÓN

Para desarrollar las aplicaciones antes mencionadas, hemos diseñado un sistema IoT con 5 nodos: 4 nodos receptores y 1 nodo de control. Los nodos receptores están equipados con un micrófono y el nodo de control implementa las funciones de sistema de procesado y control, y base de datos. Además está conectado a un altavoz para emitir las señales necesarias en caso de que sea necesario. Estas señales serán barridos en frecunecia o grabaciones de habla según el parámetros a analizar. En la Figura 1 podemos ver un esquema del sistema IoT diseñado, con los nodos receptores y el nodo de control donde se almacena también la información. Aunque tanto el broker MQTT como la base de datos pueden estar implementados en la nube, nos ha parecido pertinente incorporarlos al nodo de control también para evitar pérdida de información por una conexión deficiente o inexistente a internet.

A. *Nodos del sistema IoT*

Los 4 nodos receptores del sistema están formados por la placa Raspberry Pi 3b (RPi). Este Single Board Computer (SBC) está basado en el circuito Broadcom BCM2837, con un procesador ARM Cortex-A53 de 4 núcleos y 1,2 GHz de velocidad con unidad de procesado gráfico incluida. La placa posee WiFi incorporado (estándar IEEE 802.11 b/g/n) que es el método de conexión empleado por nuestro sistema. Cada nodo está alimentado por una batería de 3,7 V y 3800 mAh, que nos han proporcionado hasta 10 horas de autonomía en monitorización continua, lo que nos permite realizar las mediciones sin problemas, dado que las mediciones se realizan en periodos de unos pocos minutos. Es muy útil tener autonomía extra, puesto que en el caso del SII, podremos dejar al sistema monitorizando durante una clase o ponencia sin importunar a los usuarios de la sala y obtener los valores de cada posición monitorizada. Cabe destacar que se puede variar la posición de los nodos entre medidas y calcular también valores de los parámetros en tantas posiciones como deseemos, debido a que no tienen por qué ser simultáneas las medidas de la sala al completo.

Los nodos receptores que podemos ver en la Figura 2 incorporan un micrófono de condensador de bajo coste con conexión USB, con sensibilidad = -30±3 dB y un ancho de banda de 20 Hz a 16 kHz. La configuración establecida permite que cada nodo pueda adquirir y grabar el audio automáticamente durante un tiempo configurable, con resolución por muestra de 16 bits, es decir un rango dinámico de 96.33 dB con una frecuencia de muestreo de 44.1 kHz.

Como nodo de control se han utilizado diferentes plataformas, generalmente con una capacidad de cálculo más elevada que la de los nodos receptores, lo que nos ha llevado a fijar las especificaciones mínimas en un dispositivo con un procesador de doble núcleo a 2 GHz de velocidad, 4 GB de memoria RAM y 128 GB de almacenamiento disponible para que el sistema funcione de manera aceptable. El nodo de control está conectado a un altavoz auto-amplificado de 30 W y la electrónica de todos está protegida por una estructura impresa en 3D.

B. *Funcionamiento del sistema IoT*

La Figura 3 muestra el sistema de medida de parámetros de sala y su empleo en una toma de medidas. Como se ha mencionado anteriormente, el nodo de control está conectado al altavoz emisor e incorpora tanto el control como el cálculo y el almacenamiento de los valores de los parámetros de la sala a monitorizar. La sincronización de los nodos se basa en NTP, siendo su precisión temporal

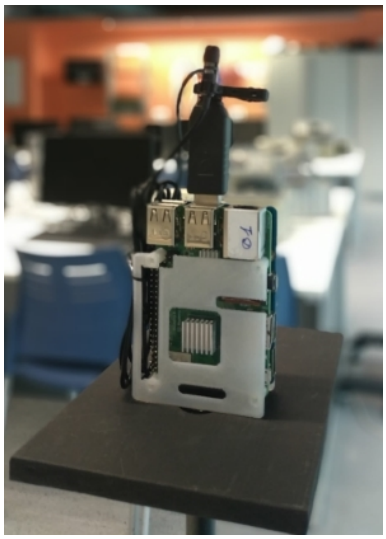


Figura 2: Nodo receptor del sistema IoT.

Figura 3: Sistema IoT instalado en un aula.

más que suficiente para la aplicación. Esta sincronización sólo es necesaria entre el nodo de control que emite sonido y los receptores, y se realiza con marcas de tiempo en el momento en que se comienza a reproducir la señal de audio. El sistema de control recae en el protocolo de comunicaciones empleado que es Message Queue Telemetry Transport (MQTT) mediante un broker MQTT implementado en el nodo de control y que gracias a un sistema de mensajería controla que se reproduzcan las señales necesarias y se monitoricen de forma síncrona. En el caso de la reproducción, se puede reproducir tanto barridos en frecuencia que nos permiten calcular la respuesta impulsiva del entorno y después realizar el cálculo de los parámetros mencionados, como diferentes señales de habla, que permitirán el cálculo directo del SII en la posición de cada nodo. Para el cálculo de los parámetros RT60, C50, C80, STI, una vez terminada la grabación, se envían las señales de audio de cada nodo al nodo de control donde se realizan los cálculos pertinentes para obtener los parámetros RT60, C50, C80, STI. En el caso de SII, no se envían los archivos de sonido ya que se calcula el parámetro en el mismo nodo, de manera que se envía solo en valor calculado en cada nodo. En todos los casos se envía un identificador del nodo para almacenar los valores calculados en la base de datos y realizar las representaciones pertinentes. Si deseamos realizar una monitorización contínua en una sala donde hay un orador presente, no se emite ninguna señal y únicamente se calcula el valor de SII en cada una de las posiciones de los nodos receptores.

Para obtener las respuestas impulsivas en las diferentes posiciones de los nodos, se realizan 5 barridos en frecuencia de 10 segundos cada uno y 5 señales Maximum Length Sequence (MLS) de 5 segundos cada una. Para el caso del análisis de SII sin un orador presente, se emplean señales de habla constante y anecóica de 10 a 3 segundos de duración. Cabe mencionar que es interesante hacer un análisis rápido, puesto que en un futuro deseamos aplicar técnicas de deep learning para predecir los parámetros en el nodo de manera más eficiente y rápida, por lo que interesa en este caso que la duración de los audios registrados sean del entorno de segundos.

El software ha sido diseñado y testeado a través de simulación en las aulas y salas de diferentes geometrías y actualmente el sistema de adquisición y cálculo está operativo, y se ha puesto a prueba en diferentes aulas de la Escuela Técnica Superior de Ingeniería (ETSE) de la Universitat de València, actualmente adaptadas a la distancia de seguridad exigida por la situación sanitaria debido al COVID-19.

IV. RESULTADOS

Una vez diseñado y probado el sistema IoT mediante simulación y señales sintéticas, hemos pasado a probar el sistema en diferentes espacios docentes de la ETSE midiendo los parámetros acústicos de sala y monitorizando durante la realización de clases para evaluar también la inteligibilidad. En la Tabla III se muestran los resultados de los parámetros medidos en cada posición de los nodos IoT.

Del conjunto de parámetros evaluados, debido a la regularidad de la sala y a que es básicamente un rectángulo sin pilares de grandes dimensiones ni formas irregulares, RT60, C50 y C80 no varían demasiado en función de la posición de medida, teniendo una media de 0.8 s de reverberación (RT60), unos 3.5 dB de claridad del habla (C50) y 6.5 dB de claridad musical (C80). Sin embargo STI y SII dependen más de la posición estudiada, siendo muy afectados (sobretodo SII) por la distancia a la fuente de sonido. La transmisión del habla (STI) baja de 0.67 a 0.64, pero la inteligibilidad cae de 0.81 a 0.71 conforme nos alejamos del hablante, lo que no representa demasiado problema en este aula, pero nos hace ver que podría afectar más en espacios más amplios. En concreto el parámetro SII nos ha parecido sumamente interesante debido a que se puede obtener a partir de señales de habla directamente y nos aporta una descripción muy útil de los problemas de percepción del habla que se pueden producir en espacios docentes. Además, nuestro sistema IoT, como hemos comentado, permite evaluar el SII en tiempo real durante la impartición de una clase, conferencia u evento, sin necesidad de interrumpir o interferir en la misma,

únicamente colocando los nodos en las ubicaciones de interés.

Cuadro III: Parámetros obtenidos con sistema IoT

Nodo	RT60 (s)	C50 (dB)	C80 (dB)	STI	SII
1	0.81	4.30	7.44	0.67	0.81
2	0.81	3.36	6.16	0.66	0.75
3	0.81	3.36	6.85	0.66	0.76
4	0.79	3.61	7.17	0.65	0.74
5	0.84	3.10	6.29	0.65	0.72
6	0.85	3.26	6.48	0.64	0.71

Además de los valores mostrados en la Tabla III que se pueden consultar en la base de datos, se ha desarrollado una aplicación que permite representar en dos y tres dimensiones la sala a tratar, junto con el parámetro que se desee en cada posición estudiada. Se ha configurado el sistema de tal forma que permita visualizar sobre un plano del entorno a monitorizar los resultados medidos de forma casi instantánea.

En la Figura 4 se muestra un ejemplo del estudio del SII en un aula de la ETSE, que tiene unas dimensiones de 12 x 8 x 3 metros cúbicos. Esta representación nos permite visualizar de una manera rápida como disminuyen el valor del parámetro SII conforme nos alejamos del hablante. De una forma más intuitiva, en la Figura 5 podemos ver un mapa de calor donde los valores más altos de SII se representan con un tono más cálido y los valores bajos en tonos cada vez más azules o fríos, junto con la escala que muestra los valores máximos y mínimos obtenidos del análisis. Esto nos permite evaluar gráficamente las distancias del hablante y las zonas en las que puede haber problemas de inteligibilidad.

Es de remarcar que el sistema dispone de 4 nodos receptores, que al ser inalámbricos se puede ubicar en sitios diferentes, lo que va a permitir evaluar la sala desde diferentes puntos y ubicaciones. Esta representación se puede obtener conectándose al nodo de control una vez almacenados los datos en el mismo, por lo que de forma automática podemos lanzar el proceso de monitorización, y tras ello recoger, guardar y procesar los resultados de forma casi instantánea, lo que en la práctica supone una gran ventaja frente a los sistemas tradicionales, menos automatizados y complejos. destacando que representa un ahorro de tiempo considerable a la hora de evaluar salas.

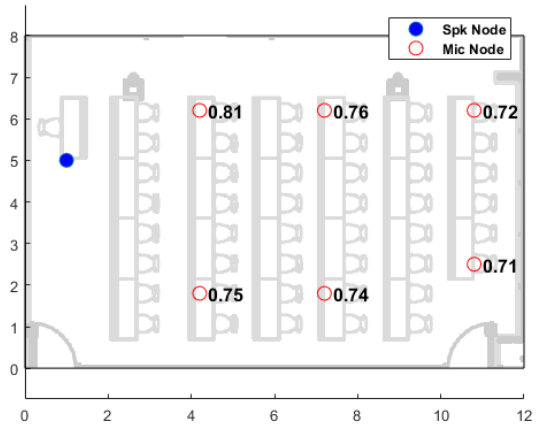


Figura 4: Representación 2D de SII en aula monitorizada.

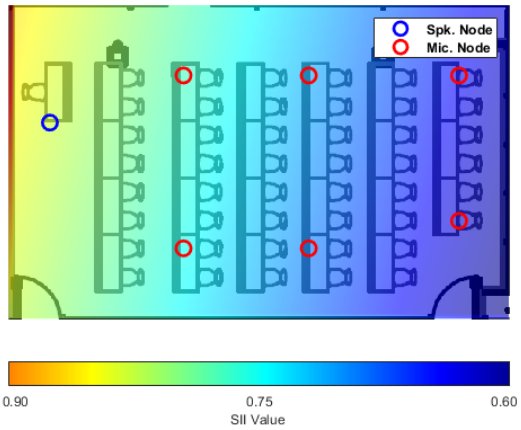


Figura 5: Mapa de calor de los valores de SII medidos.

V. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha diseñado, prototipado y desplegado un sistema IoT que permite el cálculo de los parámetros RT60, C50, C80, STI y SII, implementados en base a la estandarización y normativas asociadas referentes a medidas de los parámetros acústicos en interiores. En la época actual y debido a la distancia social necesaria que se ha implantado en los centros docentes, es necesario tener accesible un sistema como el diseñado para medir la inteligibilidad en entornos y espacios de diferente índole. Destacar de la existencia de productos propietarios que permiten realizar estudios y análisis similares, eso sí, empleando diferentes dispositivos hardware y software suministrados por separado y requiriendo por lo tanto un coste mucho más elevado si lo comparamos con el sistema propuesto en este artículo.

Resaltar que se ha reducido la complejidad del sistema de forma que el cliente final no tenga que preocuparse de los aspectos tecnológicos y nada más tenga que centrarse en las posiciones a analizar en cada entorno. Así mismo para facilitar esta tarea, los nodos de nuestro sistema son totalmente autónomos en cuanto a energía se refiere y se comunican de manera inalámbrica, por lo que facilita su despliegue y ubicación en cualquier zona. El sistema diseñado implementa una visualización de los datos de forma intuitiva, pues de forma casi instantánea nos permite visualizar el parámetro SII en las posiciones determinadas.

Además, el sistema puede almacenar los datos que se deseen en una base de datos local o en un servidor en la nube si se desea y poder visualizar estos datos en cualquier momento. Gracias a ello se pueden testear y ensayar diferentes técnicas de mejora de inteligibilidad, tanto basadas en sistemas hardware de megafonía, como software de procesado de señal, lo que va a permitir hacer un estudio simple y rápido de espacios. Todo ello permite analizar y mejorar la inteligibilidad del habla no sólo en aulas docentes, sino en cualquier tipo de recinto o

entorno donde la información oral y su transmisión sea crítica, como puede ser un quirófano, o en general una sala improvisada de reuniones.

Actualmente se está trabajando en la aceleración del cálculo mediante técnicas de deep learning y redes neuronales para permitir que éste se realice de forma más rápida y eficiente en los propios nodos IoT receptores mediante señales de habla únicamente y sin tener que calcular la respuesta impulsiva de la sala.

AGRADECIMIENTOS

Los autores agradecen a la Agencia Estatal de Investigación (AEI) y al Fondo Europeo de Desarrollo Regional (FEDER) la financiación parcial de esta investigación dentro de los proyectos BIA2016-76957-C3-1-R, RTI2018-097045-B-C21, financiados por MCIN/AEI y por "FEDER Una manera de hacer Europa", y a la ayuda BES-2017-082340, financiada por MCIN/AEI/ 10.13039/501100011033 por "FSE Invierte en tu futuro". También a la Generalitat Valenciana, por financiar la beca AEST/2020/048, BEST/2021/150, AEST/2021/016, y los proyectos GV/2020/046 y AICO/2020/154. Finalmente, a la Universitat de Valencia por la financiación de la acción especial UV-INV-AE-1544281.

REFERENCIAS

[1] Sabine, W.C. *Collected Papers on Acoustics*. Peninsula Publishing, Los Altos, CA., 1992

[2] ISO 3382-1:2009. *Acoustics–Measurement of Room Acoustic Parameters–Part 1: Performance Spaces*; International Standard Organization: Geneva, Switzerland, 2009.

[3] ISO 3382-2:2008. *Acoustics–Measurement of Room Acoustic Parameters – Part 2: Reverberation Time In Ordinary Rooms*; International Standard Organization: Geneva, Switzerland, 2008.

[4] ISO 3382-3:2012. *Acoustics–Measurement of Room Acoustic Parameters–Part 3: Open Plan Offices*; International Standard Organization: Geneva, Switzerland, 2012.

[5] ISO 9921:2004. *Ergonomics - Assessment of speech communication*; International Standard Organization: Geneva, Switzerland, 2004.

[6] ANSI/ASA S3.5:1997. *American National Standard Methods For Calculation Of The Speech Intelligibility Index*; Acoustical Society of America (ASA), 1997.

[7] Schroeder, M.R. Integrated-Impulse Method Measuring Sound Decay without Using Impulses. *J. Acoust. Soc. Am.* 1979, *66*, 497–500.

[8] Farina, A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In Proceedings of the AES 108th Convention, Audio Engineering Society, Paris, France, 19–22 February 2000; Preprint 5093.

[9] Vorlander, M.; Kob, M. Practical aspects of MLS measurements in building acoustics. *Appl. Acoust.* 1997, *52*, 239–258.

[10] Stan, G.-B.; Embrechts, J.-J.; Archambeau, D. Comparison of different impulse response measurement techniques. *J. Audio Eng. Soc.* 2002, *50*, 249–262.

[11] Mommertz, E.; Muller, S. Measuring Impulse Responses with Digitally Pre-emphasized Pseudorandom Noise Derived from Maximum-Length Sequences. *Appl. Acoust.* 1995, *44*, 195–214.

[12] Cobos M, Perez J. J., Felici S., Segura J. and Navarro J. M., Cumulative-sum-based localization of sound events in low-cost wireless acoustic sensor networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2014, *12*, 1792–1802.

[13] Alexandridis A. and Mouchtaris A., Multiple sound location estimation and counting in a Wireless Acoustic Sensor Network. In Proceedings of the 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 18–21 October 2015.

[14] Malhotra B.; Nikolaidis I.; Harms J. Distributed classification of acoustic targets in wireless audio-sensor networks. *Comput. Netw.* 2008. doi:10.1016/j.comnet.2008.05.008

[15] Duarte M.F.; Hen, H.Y. Vehicle classification in distributed sensor networks. *J. Parallel Distrib.Comput.* 2004, *64*, 826838. doi:10.1016/j.jpdc.2004.03.020

[16] Pastor-Aparicio, A.; Segura-Garcia, J.; Lopez-Ballester, J.; Felici-Castell, S.; García-Pineda, M.; Pérez-Solano, J.J. Psychoacoustic Annoyance Implementation With Wireless Acoustic Sensor Networks for Monitoring in Smart Cities. *IEEE Int. Things J.* 2020, *7*, 128–136, doi: 10.1109/JIOT.2019.2946971.

[17] van Waterschoot, T.; Moonen, M. Distributed estimation and equalization of room acoustics in a Wireless Acoustic Sensor Network. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, 27–31 August 2012.

[18] Larm, P.; Hongisto, V. Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index. *J. Acoust. Soc. Am.* 2006, *119*, 1106–1117. doi:10.1121/1.2146112

[19] Lam, C.L.C. *Improving the Speech Intelligibility in Classrooms*; Department of Mechanical Engineering, The Hong Kong Polytechnic University: Hong Kong, China, 2010.

[20] McNeer, R.R. Bennett, C.L.; Horn, D.B.; Dudaryk, R. Factors affecting acoustics and speech intelligibility in the operating room: Size matters. *Anesth Analg.* 2017, *124*, 1978–1985. doi: 10.1213/ANE.0000000000002118.

[21] Ryherd, E.E.; Moeller, M., Jr.; Hsu, T. Speech intelligibility in hospitals. *J Acoust Soc Am.* 2013, *134*, 586–595. doi: 10.1121/1.4807034.

[22] Lopez-Ballester J.; Pastor-Aparicio A.; Felici-Castell S.; Segura-Garcia J.; Cobos M. Enabling Real-Time Computation of Psycho-Acoustic Parameters in Acoustic Sensors Using Convolutional Neural Networks. *IEEE Sens. J.* 2020, *20*, 11429–11438, doi:10.1109/JSEN.2020.2995779.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Control de congestión en redes de vehículos

Ignacio Soto*, Oscar Amador†, Maria Calderon*, Manuel Urueña‡
*Departamento de Ingeniería Telemática; Universidad Carlos III de Madrid, Madrid, España
†Universidad Tecnológica de Durango, Durango, Mexico
‡Escuela Superior de Ingenieros y Tecnología, Universidad Internacional de La Rioja, La Rioja, España
isoto@it.uc3m.es, oscaramador.cele@gmail.com, maria@it.uc3m.es, manuel.uruena@unir.net

***Palabras Clave*—Dual-α, DCC, ETSI, redes de vehículos.**

El ETSI (*European Telecommunications Standards Institute*) ha desarrollado una serie de especificaciones para habilitar la comunicación entre vehículos en la banda de 5,9 GHz, que colectivamente definen una torre de protocolos denominada ITS-G5. Estas especificaciones incluyen un mecanismo de control de congestión distribuido (DCC o *Distributed Congestion Control*) [1]. Este trabajo presenta el mecanismo de DCC especificado en el ETSI, así como distintas propuestas de mejora.

La especificación en [1] recoge, en su última versión, una variante adaptativa para el mecanismo de DCC a usar en ITS-G5. Esta variante adaptativa está basada en un algoritmo, llamado LIMERIC, propuesto originalmente en [2]. LIMERIC realiza el control de congestión regulando independientemente la tasa de envío de mensajes en cada transmisor mediante un sistema de control lineal que utiliza como entrada la ocupación del canal.

Sin embargo, hasta [3] no se había realizado un estudio experimental exhaustivo de las prestaciones del mecanismo adaptativo de DCC propuesto en [1]. La conclusión principal de [3] es que este mecanismo de DCC funciona adecuadamente para cualquier rango realista de densidades de vehículos, aunque se identificaron dos debilidades que pueden afectar a sus prestaciones en ciertas situaciones.

La primera es la lentitud de convergencia cuando hay que bajar la utilización del canal, lo que puede llevar, durante periodos de tiempo transitorios pero significativos, a caídas de prestaciones o repartos injustos del medio entre diferentes transmisores. La razón de este comportamiento viene de los parámetros del algoritmo adaptativo usados en el estándar ETSI, según se razona analíticamente en [4]. En [4] se propone Dual-α, una mejora al mecanismo adaptativo de DCC especificado en el estándar ETSI, que alcanza las mismas prestaciones en régimen estacionario, pero consigue una mejor velocidad de convergencia en situaciones transitorias desde estados de alta utilización del canal.

La segunda debilidad del mecanismo de DCC del estándar se manifiesta en el envío de mensajes de concienciación (CAMs o *Cooperative Awareness Messages*). Los vehículos envían periódicamente estos mensajes para indicar a sus vecinos su situación y otros parámetros, y son la base de muchas aplicaciones de seguridad del vehículo conectado. Por su función, la información de estos mensajes debe ser lo más actual posible. La arquitectura ETSI recoge un mecanismo de realimentación que permite generar estos mensajes, como máximo, a la velocidad a la que el mecanismo de DCC va a permitir su envío. Pero, en la solución ETSI, una desincronización entre el momento de generación del CAM y el momento en el que el mecanismo de DCC permite enviar puede llevar a que el CAM espere en cola de DCC mientras aumenta la edad de la información que transporta. En [5] se propone un mecanismo, GoT (*Generate-on-Time*), que evita este problema.

AGRADECIMIENTOS

Este trabajo fue financiado por la Agencia Estatal de Investigación (AEI), proyecto PID2019-104207RB-I00/AEI/10.13039/501100011033, y por la Comunidad de Madrid a través de la línea de "Excelencia del Profesorado Universitario" del Convenio Plurianual con la UC3M (EPUC3M21), en el marco del V PRICIT (V Plan Regional de Investigación Científica e Innovación Tecnológica).

REFERENCIAS

[1] "Intelligent Transport Systems (ITS); Decentralized Congestion Control Mechanisms for Intelligent Transport Systems Operating in the 5 GHz Range; Access Layer Part", EN 102 687, V1.2.1, European Telecommunications Standards Institute, Apr. 2018.
[2] Gaurav Bansal, John B. Kenney, and Charles E. Rohrs, "LIMERIC: A linear adaptive message rate algorithm for DSRC congestion control," IEEE Trans. Veh. Technol., vol. 62, n. 9, pp. 4182-4197, Nov. 2013.
[3] Oscar Amador, Ignacio Soto, Maria Calderon, Manuel Urueña, "Experimental Evaluation of the ETSI DCC Adaptive Approach and Related Algorithms", IEEE Access, vol. 8, pp. 49798-49811, 2020.
[4] Ignacio Soto, Oscar Amador, Manuel Urueña, Maria Calderon, "Strengths and Weaknesses of the ETSI Adaptive DCC Algorithm: A Proposal for Improvement", IEEE Commun. Lett., vol. 23, n. 5, pp. 802-805, May 2019.
[5] Oscar Amador, Ignacio Soto, Manuel Urueña, Maria Calderon, "GoT: Decreasing DCC Queuing for CAM Messages", IEEE Commun. Lett., vol. 24, n. 12, pp. 2974-2978, December 2020.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Arquitectura para redes IoT orientada a la sostenibilidad medioambiental

Jorge Navarro-Ortiz, Natalia Chinchilla-Romero, Félix Delgado-Ferro, Juan J. Ramos-Munoz
Departamento de Teoría de la Señal, Telemática y Comunicaciones,
Universidad de Granada
C/ Periodista Daniel Saucedo Aranda, s/n. ETSI Informática y de Telecomunicación.
jorgenavarro@ugr.es, nataliachr@ugr.es, felixdelgado@correo.ugr.es, jjramos@ugr.es

En este trabajo se presenta una arquitectura para redes IoT orientada a la sostenibilidad medioambiental. Debido a la adecuación de sus características en términos de cobertura, potencia y soporte de un gran número de dispositivos, se ha elegido una red LoRaWAN mejorada como base de la presente propuesta. La arquitectura se completa con la virtualización mediante contenedores de las diferentes entidades de red LoRaWAN y el uso de una red definida por software para su interconexión. La publicación y suscripción a los datos medioambientales se realiza mediante el protocolo MQTT, que ha sido optimizado gracias al uso de la red SDN y al uso de recursos de *edge computing*. La arquitectura propuesta ha sido implementada mediante una red prototipo a modo de prueba de concepto.

***Palabras Clave*- IoT, LoRaWAN, MQTT, SDN**

I. INTRODUCCIÓN

El objetivo del presente trabajo es el diseño de una arquitectura de red IoT (*Internet of Things*) para la recopilación, procesado y distribución de información medioambiental. Se pretende que sea lo suficientemente flexible y potente para poder integrar, en el futuro, las diferentes soluciones desarrolladas en este ámbito.

Esta red se compone de varias partes. Por un lado, una red de acceso radio de tipo LPWAN (*Low Power Wide Area Network*) adecuada para comunicaciones de sensores masivos. Se ha elegido una red de tipo LoRaWAN (*Long Range Wide Area Network*) [1] debido a sus características adecuadas de bajo consumo, alta cobertura, fácil escalabilidad y a que utiliza una banda de frecuencias sin licencia, lo que facilita su desarrollo y reduce costes.

Por otro lado, una red troncal que contará con las diferentes entidades necesarias para este tipo de redes y para el procesado y distribución de los datos. Al usarse una red LoRaWAN para la parte radio, la red troncal incluirá un servidor de red y un servidor de aplicación, elementos necesarios en este tipo de redes. Este tipo de servidores utiliza habitualmente el protocolo MQTT (*Message Queuing Telemetry Transport*) [2] para el intercambio de información con entidades tanto internas como externas. Por ello, como se explicará en las siguientes secciones, se introducirán elementos que permitirán reducir tanto el tráfico generado como la latencia en MQTT. También se incluirá una plataforma de Inteligencia Artificial (IA) para el procesado de los datos. Todas las entidades se implementarán como funciones de red virtuales (NVFs, *Network Virtualization Functions*) de forma que se facilite su orquestación, despliegue y ejecución en nubes locales.

Por último, se utilizará una red definida por software (SDN, *Software Defined Networking*) para la comunicación entre la red de acceso radio y la red troncal. Este tipo de redes proporciona una gran flexibilidad y facilidad de desarrollo para incluir e.g. nuevos protocolos u optimizaciones sobre los existentes.

Este artículo se organiza en las siguientes secciones. La sección actual introduce los objetivos del trabajo y su contexto. La Sección II realiza una revisión del Estado del Arte. La Sección III describe la arquitectura de red IoT diseñada, exponiéndose la red prototipo realizada a modo de prueba de concepto en la Sección IV. A continuación, la Sección V presenta algunos resultados preliminares, concluyendo el artículo en la Sección VI.

II. ESTADO DEL ARTE

En este apartado se hará un breve resumen de las distintas posibilidades para desplegar una red LoRaWAN. Otros aspectos como plataformas para IoT, más centradas en el almacenamiento, distribución, procesado y visualización de datos (e.g. Google Cloud Platform, IBM Watson IoT, Amazon AWS IoT Core, Microsoft Azure, entre otros) se quedan fuera del ámbito de este trabajo.

Como ejemplo de arquitectura de red LoRaWAN de gran tamaño, cabe destacar *The Things Networks*, una red colaborativa y abierta con más de 20.000 gateways en todo el mundo. Su arquitectura [3] se compone de *bridges* que interactúan con los *gateways*, conectados a *routers* que encaminan hacia los *brokers* correspondientes (dependientes de la ubicación geográfica). Estos

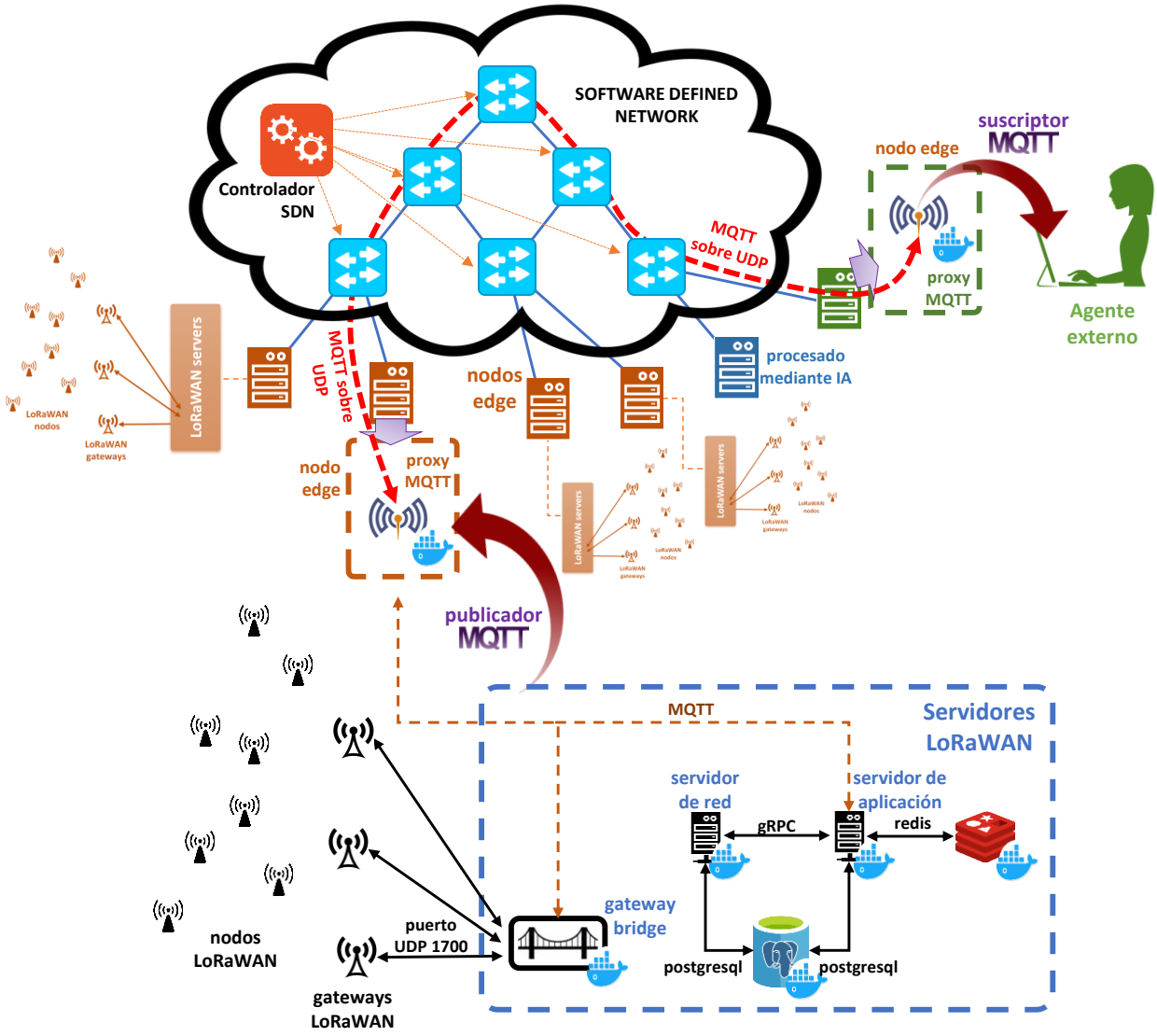


Fig. 1. Propuesta de arquitectura de red IoT para la recolección, procesado y uso de variables medioambientales.

interactúan con servidores de red y *handlers* que gestionan la comunicación con servidores de aplicación.

Respecto a plataformas LoRaWAN caben destacar Chirpstack [4], gratuita, de código abierto y que permite desplegar una red propia, y LORIOT [5], gratuita para redes pequeñas (hasta 30 nodos) pero sin posibilidad de montar una red privada propia.

III. Diseño de Arquitectura de Red IoT

La Fig. 1 muestra la arquitectura propuesta, que consiste en una red SDN que conecta *nodos edge*. Estos nodos implementarán la funcionalidad de *proxy* MQTT y/o la de procesamiento mediante IA. Estos *nodos edge* se conectarán con un nodo o *cluster* que implementará las diferentes entidades de una red LoRaWAN, es decir, el servidor de red, el servidor de aplicación y un *bridge* que permite la conexión entre el servidor de red y el *gateway*, así como las bases de datos necesarias. El *gateway* permite la conexión radio con las motas LoRaWAN, que serán los sensores y actuadores que envíen información medioambiental. En una red LoRaWAN típica también se incluiría un *broker* MQTT, que en este caso se ve reemplazado por el *proxy* MQTT.

Los *proxies* MQTT permiten la reducción del tráfico MQTT dentro de la red SDN, así como la disminución de su latencia. Para ello, actúan como *brokers* MQTT respecto a los clientes (implementando los mismos mecanismos de seguridad, e.g. usando MQTT sobre TLS/SSL), tanto suscriptores como publicadores, directamente conectados. En el caso de esta arquitectura, varias entidades de la red LoRaWAN lo utilizarán en sustitución del *broker*: *gateway bridge*, servidor de red y servidor de aplicación. Además, los datos publicados mediante MQTT podrán ser consultados por entidades externas ("*agente externo*" en la figura). Para ello, estos suscriptores MQTT se conectarán a su vez a otro *proxy* MQTT que actuará como su *broker* e intercambiará la información necesaria con el otro *proxy* usando UDP. La ventaja de usar UDP es que se reduce la latencia (varios *Round-Trip Times* (RTT) en caso de usar TCP) y nos permitirá en un futuro enviarlo por multidifusión (véanse los trabajos futuros en las conclusiones). Además, si bien no se ha incluido en la prueba de concepto, se puede utilizar la solución dada en [6] para garantizar la fiabilidad de UDP sobre SDN.

Concretamente, los *proxies* reenviarán los mensajes de suscripción (*Subscribe*) –cuando aparezca un nuevo tópico que no tuviese suscriptores previos en ese *proxy*–, publicación (*Publish*) –para todos los mensajes publicados en tópicos que tengan algún suscriptor en el *proxy* destino– y desconexión (*Disconnect*) –cuando se desconecte el último suscriptor a ese tópico en ese proxy–. De esta manera, los *proxies* tendrán la información necesaria para reenviarse los mensajes de tópicos activos.

IV. Prueba de Concepto

Como prueba de concepto, se ha implementado una red SDN utilizando *mininet* [7] siguiendo una topología en árbol con 3 *switches*, i.e. un *switch* raíz (*s1*) y dos *switches* hoja (*s2* y *s3*), que conectan 4 nodos *edge* (*h1* y *h2* conectados a *s2*, y *h3 y h4* conectados a *s3*). Los nodos *h1* y *h4* ejecutan sendos *proxies* MQTT que han sido programados utilizando *Scapy* [8]. El nodo *h1* utiliza un interfaz de red real del PC en el que se ejecuta, de manera que permite conectar directamente con la red troncal LoRaWAN. En este caso, por sencillez, se ha optado por utilizar contenedores, orquestados usando Kubernetes [9], que ejecutan las diferentes entidades de la plataforma Chirpstack (*gateway bridge* y servidores de red y aplicación LoRaWAN) en el propio *gateway* LoRaWAN, de forma similar a [10]. Dicho *gateway* es un Lite Gateway [11] de IMST, que utiliza una Raspberry Pi junto con un concentrador LoRaWAN iC880A. Se dispone de motas FiPy [12] con placas de expansión PySense. Como controlador SDN se ha utilizado RYU [13], que ejecuta un *learning switch* para OpenFlow v1.3 [14]. Los datos enviados por MQTT son almacenados en una base de datos InfluxDB por un *script* Python. Estos datos finalmente son visualizados utilizando para ello Grafana [15]. Como demostración de su funcionamiento, la Fig. 2 muestra las trazas de todos los equipos cuando *h2* se conecta a *h1* como suscriptor (usando `mosquitto_sub`) y *h4* manda un mensaje "*message1*" como publicador al tópico "topic1" (usando `mosquitto_pub`). Tal como se muestra en la traza *Wireshark*, tomada en *h1*, *h1* responde a la petición de suscripción de *h2* comportándose como haría un *broker*. Se pueden observar los mensajes MQTT enviados (*Connect*, *ConnAck*, *Subscribe*, *SubAck*). Concretamente, se puede ver un mensaje UDP después del mensaje *Subscribe Request*, que es el reenvío de dicho mensaje al otro *proxy* usando UDP en vez de TCP. Posteriormente, *h3* manda el mensaje MQTT *Publish Request* a *h4*, que lo reenvía a *h1* usando UDP (también visible en la traza *Wireshark*) y este lo publica de forma que lo recibe *h2*, que lo muestra en su consola.

V. Resultados Preliminares

A modo de resultados preliminares, la Fig. 3 muestra parte de los datos visualizados en Grafana. Se muestran, a modo de ejemplo, valores de temperatura, humedad y luz instantáneos y gráficas con los valores del último día. En los *dashboards* implementados también se visualizan gráficas de la última hora y con los valores mínimo, medio y máximo de estas métricas para el último día por horas, y para la última semana y mes por días.

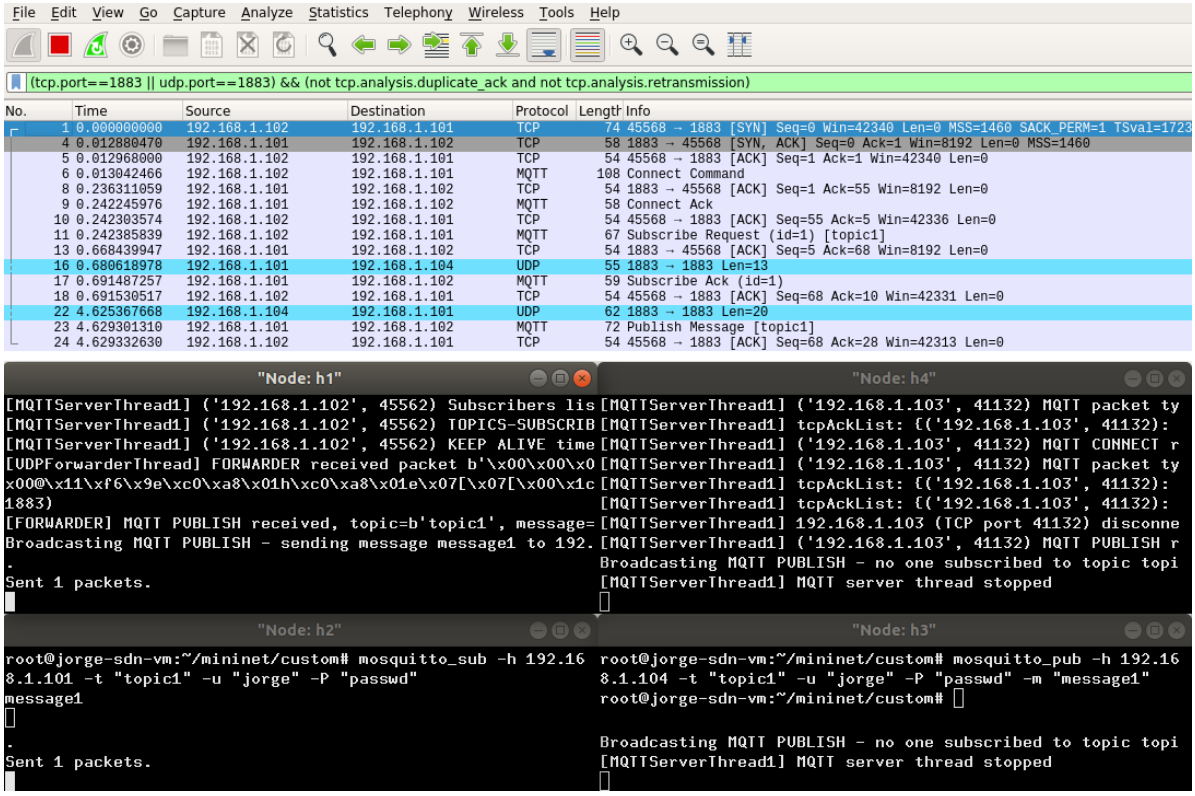


Fig. 2. Funcionamiento de proxies MQTT en los *edge switches*.

Fig. 3. Ejemplo de visualización de variables medioambientales.

También se ha comprobado la reducción de latencia y tráfico MQTT gracias al uso de UDP en la red SDN. Respecto al tráfico, una publicación MQTT normal requiere 12 mensajes (establecimiento de conexión TCP, Connect/ConnAck, Publish, Disconnect, finalización de conexión, ACKs de TCP), i.e. unos 550 bytes suponiendo cabeceras IP y TCP sin opciones y para tópicos y mensajes de tamaño pequeño (12 bytes entre ambos). Al utilizar nuestra solución, solo el mensaje *Publish* atravesaría la red SDN con unos 52 bytes más las longitudes del tópico y el mensaje (64 bytes en el ejemplo), lo que implica una reducción de tráfico de un 88%. Esta reducción en mensajes implica una menor latencia (desde que se publica el dato hasta que se recibe), efecto más apreciable para valores de RTT elevados, tal como muestra la Fig. 4. Con el objetivo de que la comparativa sea justa, el *broker* será una modificación de nuestro *proxy* realizado con Scapy. En media, la latencia se reduce aprox. un 80%.

VI. Conclusiones

En este artículo se ha presentado el diseño de una arquitectura de red IoT orientada a la recopilación, procesado y visualización de variables medioambientales. Además de realizar una prueba de concepto, el prototipo incluye una modificación de MQTT que divide la conexión TCP en partes, usando UDP para el intercambio de mensajes de MQTT entre *proxies* a través de la red SDN. Este prototipo ha permitido mostrar algunos resultados preliminares como la reducción de latencia y tráfico MQTT y la visualización de datos medioambientales.

Para mejorar la arquitectura de red propuesta, los autores prevén los siguientes trabajos:

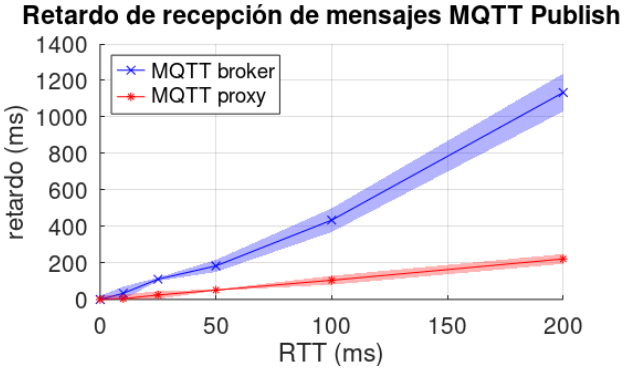


Fig. 4. Resultados de latencia y tráfico generado MQTT.

- Mejora de la capacidad de la red LoRaWAN. Este trabajo ya ha sido realizado, consiguiéndose aumentar la capacidad un 95% con condiciones radio ideales y un 40% con modelos de propagación realistas [16].
- Mejora de la tasa de datos de los dispositivos LoRaWAN. Esta mejora permitiría utilizar el mismo dispositivo para enviar fotografías y vídeo de baja resolución en tiempo real que podrían usarse para e.g. confirmar situaciones de alarma.
- Uso de una plataforma de Inteligencia Artificial o *Machine Learning* para el procesado de datos para, por ejemplo, predecir series temporales o eventos futuros.
- Dado el gran tamaño de una red IoT de este tipo, se podría optimizar el envío de tráfico MQTT utilizando un protocolo de multidifusión dentro de la red SDN.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Agencia Andaluza del Conocimiento (proyecto A-TIC-241-UGR18), el Ministerio de Economía y Competitividad (proyecto TEC2016-76795-C6-4-R) y el proyecto H2020 5G-CLARITY (Grant No. 871428).

Referencias

[1] LoRaWAN Specification v1.1, LoRa Alliance, 2017. Disponible en https://lora-alliance.org/resource_hub/lorawan-specification-v1-1/ (visitado el 30/5/2021).

[2] MQTT v3.1 Protocol Specification, IBM and Eurotech, 2010. Dispoinble en https://public.dhe.ibm.com/software/dw/webservices/ws-mqtt/mqtt-v3r1.html (visitado el 30/5/2021).

[3] The Things Network – Network Architecture. Disponible en https://www.thethingsnetwork.org/docs/network/architecture/ (visitada el 30/5/2021).

[4] Chirpstack, Open-Source LoRaWAN Network Server Stack. Disponible en https://www.chirpstack.io/ (visitada el 30/5/2021).

[5] LORIOT – Connecting the Internet of Things. Disponible en https://www.loriot.io/ (visitada el 30/5/2021).

[6] M. Wang, L. Chen, P. Chi and C. Lei, "SDUDP: A Reliable UDP-Based Transmission Protocol Over SDN," in IEEE Access, vol. 5, pp. 5904-5916, 2017, doi: 10.1109/ACCESS.2017.2693376.

[7] Mininet – An Instant Virtual Network on your Laptop. Disponible en http://mininet.org/ (visitado el 30/5/2021).

[8] Scapy – Packet Crafting for Python2 and Python3. Disponible en https://scapy.net/ (visitado el 30/5/2021).

[9] Kubernetes – Production-Grade Container Orchestration. Disponible en https://kubernetes.io/ (visitado el 30/5/2021).

[10] J. Navarro-Ortiz, J. J. Ramos-Munoz, J. M. Lopez-Soler, C. Cervello-Pastor, M. Catalan, "A LoRaWAN Testbed Design for Supporting Critical Situations: Prototype and Evaluation", Wireless Communications and Mobile Computing, vol. 2019, DOI: 10.1155/2019/1684906

[11] Lite Gateway – Demonstration Platform for LoRa Technology, IMST. Disponible en https://wireless-solutions.de/products/lora-solutions-by-imst/development-tools/lite-gateway/ (visitado el 30/5/2021).

[12] FiPy Development Board, Pycom. Disponible en https://pycom.io/product/fipy/ (visitado el 30/5/2021).

[13] RYU – Component-Based Software Defined Networking Framework. Disponible en https://ryu-sdn.org/ (visitado el 30/5/2021).

[14] OpenFlow Switch Specification, Version 1.3.0. Open Networking Foundation, 2014. Disponible en https://opennetworking.org/wp-content/uploads/2014/10/openflow-spec-v1.3.0.pdf (visitado el 30/5/2021).

[15] Grafana – Your Observability Wherever You Need It. GrafanaLabs. Disponible en https://grafana.com/ (visitado el 30/5/2021).

[16] Natalia Chinchilla-Romero, Jorge Navarro-Ortiz, Pablo Muñoz, Pablo Ameigeiras, "Collision Avoidance Resource Allocation for LoRaWAN", Sensors, 21 (4), 2021. DOI: 10.3390/s21041218.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Predictores de energía recolectada en redes de sensores: sencillez y eficiencia

S. Herrería Alonso, A. Suárez González, M. Rodríguez Pérez, R. Rodríguez Rubio, C. López García
atlanTTic Research Center,
Universidade de Vigo,
Rúa Maxwell s/n, Vigo 36310.
{sha,asuarez,miguel,rrubio,candido}@det.uvigo.es

La luz solar y el viento son dos de las fuentes de energía ambientales más accesibles para el posible abastecimiento de baterías recargables en las redes de sensores inalámbricas. Pero, aunque virtualmente ilimitadas, su aporte energético también es impredecible ya que ambas sufren variaciones significativas debido a unas condiciones climáticas variables (estación del año, momento del día, emplazamiento geográfico...). Es por ello que resulta de utilidad que los elementos de la red (sensores) usen predictores que les permitan una adaptación efectiva de su gasto energético a la dinámica prevista de captación de energía. Resulta asimismo de interés que, en el caso de dispositivos con capacidades limitadas (característica habitual en las redes de sensores), estos predictores sean de baja complejidad para no despilfarrar la energía disponible en sus cálculos. En esta ponencia presentamos brevemente los predictores *Ángulo de altura solar* y *ARIMA adaptativo*, comparándolos con predictores propuestos en el mismo campo.

***Palabras Clave*—gestión de energía, energía recolectada, predicción de energía, energía solar, energía eólica**

I. INTRODUCCIÓN

Siendo la luz solar y el viento las fuentes de energía naturales más atractivas para suministrar energía dinámicamente a los elementos en una red de sensores, resulta de interés el disponer de predictores sencillos que permitan al sensor una adaptación efectiva de su gasto energético a la dinámica de su reabastecimiento/captación de energía.

En esta ponencia presentamos dos métodos de predicción de captación de energía, *Ángulo de altura solar* y *ARIMA adaptativo*, diseñados específicamente para el caso solar y eólico, respectivamente. Asimismo mostramos una comparativa con respecto a dos predictores previamente propuestos en la literatura para el mismo fin, Pro-Energy [1] y D-WCMA/UD-WCMA [2], utilizando como métrica de comparación el error absoluto medio (*Mean Absolute Error*)

$$\mathrm{MAE} = \frac{\sum |E_{[n,n+h]} - \widehat{E}_{[n,n+h]}|}{\text{numero de predicciones}}\,, \qquad (1)$$

con $E_{[n,n+h]}$ y $\widehat{E}_{[n,n+h]}$ la muestra actual y la predicha para intervalos n a $n+h$ respectivamente.

II. ÁNGULO DE ALTURA SOLAR

En [3] presentamos un sencillo predictor de la energía solar disponible en función del ángulo de altura solar θ' (AAS), tal y como se muestra en la figura 1, y del último dato de energía recolectada. Y evaluamos también una implementación simplificada de nuestro algoritmo, utilizando la función seno (función de los tiempos de salida y puesta del sol, y ángulo de incidencia al mediodía solar) en vez del valor exacto original, de cálculo más pesado.

$$\theta_t \approx \theta_{\mathrm{noon}} \sin\left(\pi \frac{t - t_{\mathrm{rise}}}{t_{\mathrm{set}} - t_{\mathrm{rise}}}\right), \quad t_{\mathrm{rise}} \le t \le t_{\mathrm{set}}\,, \qquad (2)$$

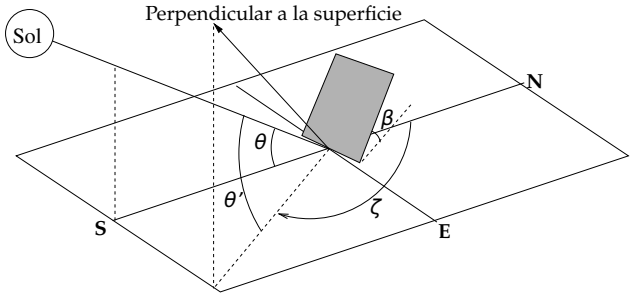


Fig. 1. Ángulo de incidencia solar sobre una superficie inclinada

En la figura 2 se muestra su mejor desempeño en el corto plazo (hasta 1 hora) —incluso considerando la implementación simplificada (AAS-sen)— para el caso de una traza del NREL [4]; y con menor carga computacional

en comparación con predictores más elaborados (Pro-Energy y D-WCMA) que sólo tienen en cuenta el historial de datos de energía generada,

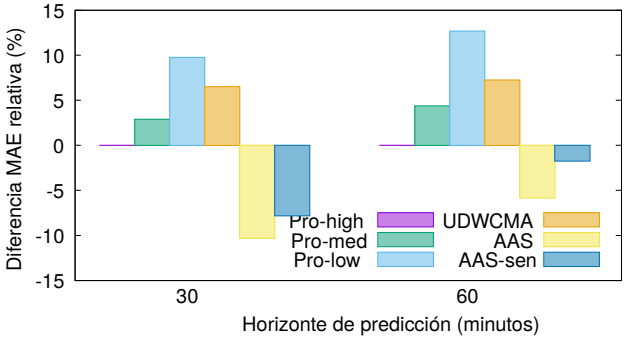


Fig. 2. *Mean Absolute Error* respecto a Pro-High

III. ARIMA ADAPTATIVO

En [5] presentamos un sencillo predictor de energía del viento, seleccionando en tiempo de ejecución bien un modelo ARIMA(1,1,1)

$$\bigtriangledown s_n - \varphi_1 \bigtriangledown s_{n-1} = \varepsilon_n - \theta_1 \varepsilon_{n-1}, \tag{3}$$

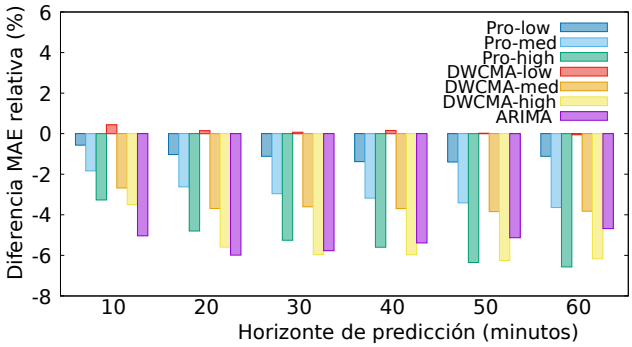
bien uno ARIMA(0,1,2) —si el anterior resulta en un estimador inestable—

$$\bigtriangledown s_n = \varepsilon_n - \theta_1 \varepsilon_{n-1} - \theta_2 \varepsilon_{n-2}. \tag{4}$$

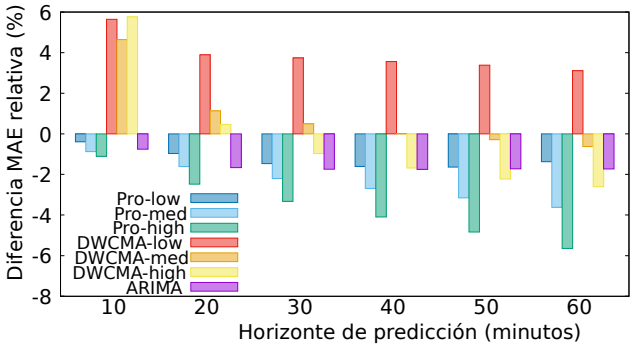
En la figura 3 se compara en el corto plazo su poder predictivo con los predictores Pro-Energy y D-WCMA para dos trazas. En los resultados para la traza ORNL 2018 [4] de la figura 3(a) —donde el ARIMA adaptativo se estabiliza rápidamente en el modo ARIMA(1,1,1)— se aprecia cómo sólo al aumentar el horizonte de predicción hacia la media hora consiguen ser superiores los predictores de mayor carga computacional (Pro-Energy-high y DWCMA-high), mientras los de carga intermedia y baja siguen dando estimaciones peores. En la traza SRRL 2019 [6], sin embargo, se aprecia como, si bien es superado por Pro-Energy-high ya desde el horizonte temporal más corto, el ARIMA adaptativo —estabilizado rápidamente en el modo ARIMA(0,1,2)— da resultados parejos a los de media y baja carga hasta el horizonte temporal de media hora.

IV. CONCLUSIONES

En el ámbito de la red de sensores, donde los elementos de la red tendrán limitaciones inherentes a su diseño y despliegue, tanto en potencia de cálculo como, más general, en disponibilidad de energía, el aumento de complejidad en los predictores de captación de energía del ambiente puede no estar justificado. En los casos particulares de energía solar y energía eólica hemos presentado sendos predictores de complejidad baja a muy baja, capaces de obtener resultados tan precisos como predictores previos de mayor complejidad, sobre todo al corto plazo, en horizonte temporal de media hora.



(a) Traza ORNL 2018



(b) Traza SRRL 2019

Fig. 3. *Mean Absolute Error* respecto a estimador persistente

AGRADECIMIENTOS

Esta publicación es parte del proyecto de I+D+i / PID2020-113240RB-I00, financiado por MCIN/ AEI/10.13039/501100011033/.

REFERENCIAS

[1] A. Cammarano, C. Petrioli, and D. Spenza, "Online energy harvesting prediction in environmentally powered wireless sensor networks," *IEEE Sensors Journal*, vol. 16, no. 17, pp. 6793–6804, sep 2016.

[2] A. H. Dehwah, S. Elmetennani, and C. Claudel, "UD-WCMA: An energy estimation and forecast scheme for solar powered wireless sensor networks," *Journal of Network and Computer Applications*, vol. 90, pp. 17–25, jul 2017.

[3] S. Herrería-Alonso, A. Suárez-González, M. Rodríguez-Pérez, R. F. Rodríguez-Rubio, and C. López-García, "A solar altitude angle model for efficient solar energy predictions," *Sensors*, vol. 20, no. 5, p. 1391, 2020.

[4] C. Maxey and A. Andreas, "Oak Ridge National Laboratory (ORNL). Rotating Shadowband Radiometer (RSR)," NREL Report No. DA-5500-56512, Oak Ridge, TN, USA, 2007, http://dx.doi.org/10.5439/1052553.

[5] S. Herrería-Alonso, A. Suárez-González, M. Rodríguez-Pérez, R. F. Rodríguez-Rubio, and C. López-García, "Efficient wind speed forecasting for resource-constrained sensor devices," *Sensors*, vol. 21, no. 3, p. 983, 2021.

[6] A. Andreas and T. Stoffel, "Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS)," NREL Report No. DA-5500-56488, Golden, CO, USA, 1981, http://dx.doi.org/10.5439/1052221.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

WBAN energéticamente eficientes y seguras mediante blockchain

J. Ramis Bibiloni[1,2], M. M. Payeras Capellà[1], L. Carrasco Martorell[1,2], and M. Mut Puigserver[1]

[1]Universitat de les Illes Balears, UIB, 07122, Palma, España
[2]Institut d'Investigació Sanitària Illes Balears, IdISBa, 07010 Palma, España.
{jaume.ramis, mpayeras, loren.carrasco, macia.mut} @uib.es

Resumen

Uno de los principales avances de la eHealth se fundamenta en la capacidad de las redes de área corporal inalámbricas (WBAN, Wireless Body Area Network) para capturar los datos biomédicos, almacenarlos y transmitirlos al destino que corresponda, de forma imperceptible y transparente. Las limitadas capacidades de los sensores precisan de protocolos de comunicaciones que reduzcan el consumo energético mediante una planificación eficiente de las transmisiones. Además, en la gestión de datos médicos, no solo se debe garantizar que los datos son auténticos, inalterados y generados por actores autorizados, sino que el acceso a estos datos debe ser restringido. Por ello se desarrollarán las herramientas para gestionar los datos médicos recolectados de forma segura y privada, así como automatizar diversas aplicaciones para su tratamiento, mediante el uso de la tecnología blockchain y los smart contracts.

***Palabras Clave*—WBAN, Blockchain, Eficiencia Energética, Tratamiento Seguro de Datos, Telemedicina, IoMT**

I. INTRODUCCIÓN

Las previsiones para el año 2030 en España estiman que unos doce millones de personas tendrán más de 65 años y que unos dos millones superarán los 80. Muchos de ellos vivirán solos y, además, padecerán enfermedades crónicas. Será necesario un seguimiento de su estado de salud y de su evolución, lo cual requerirá la monitorización de sus constantes vitales y el control estricto de las pautas de medicación. Debe cambiarse el paradigma asistencial actual, incorporando el uso de las Tecnologías de la Información y las Comunicaciones (TIC) para el aprovechamiento eficiente de los limitados recursos sociosanitarios disponibles. Se trata del concepto de eHealth, eSalud en español, que la Organización Mundial de la Salud (OMS) define como el apoyo que la utilización de las TIC ofrece a la salud y a los ámbitos relacionados con ella.

En un mundo de servicios digitalizados, con múltiples amenazas a la integridad, autenticidad y confidencialidad de los datos, se precisan fuertes medidas de seguridad para su protección. En el caso de datos relativos a la salud, estos requerimientos se hacen todavía más patentes dado que son considerados datos de nivel alto de seguridad en el Reglamento General de Protección de Datos (RGPD). Por tanto, las aplicaciones que traten con datos relativos a la salud, como es el caso del trabajo de investigación en curso, deben tener en cuenta sus altos niveles de privacidad.

Internet of Medical Things (IoMT) es un componente esencial en los modernos sistemas de gestión de la salud. Los dos grandes retos que debe afrontar la IoMT son la seguridad y la eficiencia energética. La estructura de un sistema IoMT está formada por tres niveles: Percepción, Red y Aplicación.

- El primer nivel, el de percepción, está formado por dispositivos sensores. Con el fin de favorecer la comodidad de los pacientes monitorizados, dichos dispositivos deben ser mayoritariamente inalámbricos, de tamaño reducido y poco invasivos.
- El nivel de red es el sistema inalámbrico o cableado y el middleware asociado que procesa y comunica las entradas adquiridas por los sensores del nivel de percepción. El diseño de los protocolos de comunicación juega un papel crucial en la minimización del consumo energético de los dispositivos sensores, prolongando así la vida útil de sus baterías, aspecto crítico en estas redes, en las que la intervención exterior debe reducirse al máximo.
- El nivel de aplicación incorpora los sistemas de gestión de datos médicos e historiales clínicos electrónicos (EHR, Electronic Health Record) para proporcionarlos a los diferentes actores involucrados y satisfacer las necesidades individuales. La privacidad del paciente y la seguridad de los datos son componentes esenciales en el sistema IoMT. Los datos necesitan ser adquiridos, acumulados y enviados de forma segura a las personas autorizadas.

II. ESTADO DEL ARTE

Dos de los retos más importantes a los que se enfrenta la IoMT en general y las WBAN en particular son la seguridad y la eficiencia energética [1, 2]. Existen actualmente dos estándares para las WBAN, el estándar IEEE 802.15.6 de 2012 y el más moderno ETSI SmartBAN (Smart Body Area Network) de 2015. Análisis comparativos [3, 4] entre ambas propuestas muestran que el estándar europeo presenta unos tiempos de conexión y un consumo de energía menores, lo cual nos ha impulsado a investigar las capacidades de este estándar.

Las especificaciones de la capa de control de acceso al medio (MAC, Medium Access Control) pueden influir significativamente en el consumo de energía del sensor. Aunque la mayoría del tráfico generado por los sensores corporales es periódico, existen pocos estudios que investiguen el acceso planificado en las WBAN [4–8] y todavía menos que basen su estudio en el estándar SmartBAN [4, 8]. Es por ello que nuestro trabajo se centra en la transmisión de datos periódicos con un acceso programado en el contexto de este estándar.

Los trabajos [6, 7] presentan protocolos MAC para redes WBAN con una topología en estrella que no se ajusta a ninguno de los estándares existentes. Los autores en [5] adaptan la tasa y la potencia de transmisión del sensor, minimizando la energía consumida y manteniendo la calidad de servicio (QoS, Quality of Service). Su principal limitación es que requiere la resolución de un problema de optimización complejo en el nodo coordinador o hub. En [4] se presenta un protocolo para el acceso programado en un entorno SmartBAN para minimizar el retardo, considerando la energía consumida por los sensores. La reducción del retardo se produce a costa de un mayor coste energético. En [8] se presenta una propuesta de modificación al estándar SmartBAN proponiendo una duración de la trama TDMA variable, adaptándose a los requerimientos de tráfico del sensor y obteniendo mejoras importantes de prestaciones en throughput y ahorro energético. Sin embargo, esta estrategia degrada rápidamente la probabilidad de errores (BER, Bit Error Rate) en las transmisiones a través del canal WBAN.

Por otra parte, cabe indicar que ninguno de los artículos mencionados considera la recolección de energía por parte de los sensores. En este contexto, el reto es la gestión de la energía de manera que el balance entre energía recolectada y energía utilizada se mantenga equilibrado [9]. Dado que existe muy poca bibliografía que investigue este problema [10], esta es una de las líneas más prometedoras de estudio dentro del presente trabajo en curso.

Un último aspecto a considerar es la incorporación de información sobre el estado de salud del paciente para modular el comportamiento de los protocolos de comunicación. En [11] se presenta uno de los pocos trabajos en este campo, donde el estado del paciente se utiliza para modular el número de muestras que toma y transmite el sensor. En este proyecto pretendemos llegar más lejos. Estamos investigando la incorporación de este tipo de información a todos los niveles en la configuración de los parámetros de la red y de los protocolos de comunicación.

La presente propuesta incluye el uso de blockchain para la gestión de los datos médicos. Esta tecnología ha sido revolucionaria en los últimos años por su enorme perspectiva de impacto en el campo de las plataformas seguras para compartir datos.

Otras aplicaciones de la tecnología blockchain se relacionan con el Internet de las cosas (IoT). En el contexto de IoMT y, más específicamente en el de las WBAN, el hub debe gestionar los datos de forma segura y privada, generando un mecanismo de acceso para que los datos puedan ser analizados por los servicios médicos correspondientes. El modelo de comunicación completo de un sistema de telemedicina incluye diferentes partes, como pacientes, proveedor de la red, proveedores de servicios de salud, médicos de emergencia, médicos, enfermeras, etc. Por lo tanto, los datos recogidos por la WBAN de un paciente requieren acceso desde múltiples fuentes. La difusión de datos o el acceso a diferentes intervalos del proceso de diagnóstico pueden considerarse bloques de transacción en el contexto de la tecnología blockchain.

La idea de integrar blockchain y WBAN es reciente. Entre los trabajos pioneros en esta integración encontramos [12] donde se utilizó una tecnología Ethereum blockchain para crear una plataforma estructural que permitiera a las partes rastrear la historia pasada de los pacientes. En [13] se propuso un modelo estructural para la incorporación de blockchain en las WBAN. El mismo artículo presenta un estudio sobre los retos en los que es necesario investigar para una integración exitosa de blockchain y WBAN.

Desde 2019 se han realizado algunas propuestas referentes al uso de smart contracts para la monitorización de pacientes [14–16]. En este último se afirma que la integración de IoT y blockchain es una alternativa razonable a los sistemas de salud basados en IoT centralizados. [17] presenta una arquitectura para la integración de WBAN y blockchain al tiempo que menciona algunos de los retos a los que se enfrenta el campo de las aplicaciones médicas.

En los primeros artículos de 2021 referentes al uso de blockchain en el tratamiento de datos médicos [18, 19] se analiza cómo el uso de blockchain en la gestión de datos médicos puede conducir a estimular la innovación, describiendo las principales ventajas de adoptar la tecnología blockchain en la industria de la salud. [19] presenta casos de proyecto en curso para mostrar la practicabilidad de blockchain en diferentes sectores del sistema sanitario. Finalmente, estos últimos artículos muestran cómo los servicios de telemedicina basados en blockchain pueden ser útiles para evitar la diseminación de la COVID.

III. DESCRIPCIÓN DEL TRABAJO, OBJETIVOS Y RESULTADOS ALCANZADOS.

Los objetivos científicos y tecnológicos del presente trabajo de investigación así como los resultados ya obtenidos se describirán a continuación.

Se pretende aportar soluciones a dos de los grandes retos a los que se enfrenta la IoMT para convertirse en una tecnología clave en la evolución del sistema sanitario

y del cuidado de la salud de las personas: por un lado la independencia energética de los sensores a través de protocolos con una alta eficiencia energética y sistemas de recolección de energía y, por otra, de la seguridad, protección y correcta gestión de los datos generados. Una característica a destacar de la presente investigación es el hecho de abarcar los diferentes niveles de un sistema IoMT: los niveles de percepción y transmisión, abordados en el primer gran objetivo, y el nivel de aplicación, tratado en el segundo. Al trabajar conjuntamente en los diferentes niveles es de esperar que se produzcan realimentaciones entre los objetivos que permitan, en la etapa final del proyecto, determinar aquellos aspectos transversales que requieran una aproximación multinivel.

El primer objetivo de este proyecto se centra en incrementar la autonomía energética de las WBAN a través de la optimización energética de los protocolos de comunicación. Tras revisar en profundidad las soluciones ya existentes, se ha desarrollado el diseño y análisis de protocolos de comunicaciones específicos para el acceso planificado del estándar ETSI SmartBAN, dando como resultado las publicaciones [20, 21]. Se trata de protocolos que consideran el estado del canal y los requisitos de QoS mínimos del tráfico. Los algoritmos actualmente ya desarrollados presentan un reducido coste computacional, y cuentan con el sensor para la toma de decisiones, permitiendo una rápida respuesta a los cambios de entorno. En este contexto, no se descartan propuestas para el acceso asíncrono provocado por avisos y señales de emergencia. El último logro alcanzado ha consistido en la incorporación de la recolección de energía en los sensores, logrando una mejora significativa del tiempo de vida de las baterías, cuyos resultados han sido recientemente presentados en [22]. Se ha demostrado que el mantenimiento del balance energético permite alcanzar niveles de QoS y seguridad de los datos por encima de los mínimos requeridos si el nivel de la batería y el flujo de recolección de energía son suficientes. Estos prometedores resultados alentan a seguir ahondando en esta línea, incorporando la información sobre la disponibilidad de energía (estado de las baterías y posibilidad de recolección) en el diseño de los protocolos.

El trabajo que actualmente se está llevando a cabo consiste en la adaptación de los protocolos ya desarrollados al nivel de carga de la batería del sensor. Este objetivo deberá conseguirse sin comprometer los requisitos de QoS específicos de los datos biomédicos objeto de la monitorización, con especial atención al retardo.

Se está estudiando la incorporación de información sobre el estado del paciente para modular el comportamiento de los protocolos de comunicación. En este sentido, se están analizando protocolos de muestreo adaptativo de los datos biomédicos en los sensores con el objetivo de reducir el número de medidas realizadas y minimizar las transmisiones por parte de los sensores, satisfaciendo siempre los tiempos màximos de demora asumibles.

Tanto en el caso de las posibilidades ofrecidas en el mantenimiento del balance energético como en la incorporación de datos de estado del paciente, se estudiará también la posibilidad de que este tipo de información se introduzca como elementos que afecten y sean considerados dentro del esquema de seguridad que se analizará en el segundo gran bloque del proyecto.

El segundo gran objetivo de este trabajo de investigación pretende dotar a un sistema de telemedicina de las herramientas para gestionar los datos médicos recolectados por la WBAN de forma segura y privada, de manera que los datos recogidos por las fuentes no sean alterados ni falsificados, al tiempo que se disponga de un mecanismo de auditoría público y a tiempo real, así como automatizar diversas aplicaciones para su tratamiento. Con la solución que pretende aportar el trabajo se sentarían las bases para la implementación de un sistema de telemedicina extremo a extremo que podría ser utilizado de dos formas independientes. Por una parte se podría monitorizar la salud de pacientes concretos creando un escenario con diferentes roles para distintos actores como pacientes, personal sanitario, entidades de salud, ... mientras que por otra parte se dispondría de una herramienta de gestión de datos agregados y anonimizados que permitiría el análisis de datos médicos masivos, enfocada al estudio de determinadas enfermedades o situaciones excepcionales relacionadas con la salud.

Los primeros objetivos alcanzados por el equipo de investigación sobre el diseño de aplicaciones que hacen uso de blockchain son protocolos de entrega certificada de datos, notificaciones certificadas y firma electrónica de contratos, servicios en los que los miembros del equipo investigador han trabajado en los últimos años diseñando y publicado diversos protocolos así como presentando una patente sobre uno de ellos [23–26].

La entrega registrada de datos requiere un intercambio equitativo de valores: unos datos y una prueba de no repudio de origen a cambio de una evidencia de recepción. En [23] se presentó el primer sistema basado en blockchain para este servicio. En este artículo se propusieron dos soluciones que permiten enviar notificaciones certificadas cuando se requiere confidencialidad y cuando es necesario registrar el contenido de la notificación, respectivamente. Estas entregas de datos pueden realizarse sobre blockchain como aparece en [26]. También se ha demostrado como la entrega certificada de datos mediante blockchain puede conseguir la característica de confidencialidad [25]. Esta característica puede ser aplicable a otras operaciones, como aparece en [24].

En el trabajo se continuará con el desarrolo del sistema de almacenamiento seguro de la información recogida por la WBAN, pudiéndose tratar de datos aislados o de historiales clínicos electrónicos completos. Los datos pueden ser procesados tanto como datos de un paciente concreto que requiere atención individualizada como ser tratados como datos agregados anonimizados en estudios que requieran bases de datos médicos. Tanto en uno como en otro caso los datos deben ser procesados por aplicaciones que protejan la privacidad del usuario, protejan la identidad, limiten la compartición de la información personal, garanticen que los usuarios puedan dar permisos

selectivos a la hora de compartir los datos y garanticen el origen de los mismos y su integridad/inmutabilidad. Entre los resultados alcanzados se encuentra el diseño y la implementación de una aplicación de gestión de los datos recogidos en la WBAN para que puedan ser manejados cumpliendo los requerimientos listados anteriormente, es decir, un tratamiento seguro, automatizado y personalizado de los datos mediante la programación de smart contracts.

Entre los resultados esperados se pretende añadir la capacidad de disparar eventos relacionados con las informaciones recopiladas así como la captación de mensajes off-chain por parte de los smart contracts sobre los dispositivos. El estado de las baterías permite emitir los correspondiente eventos en función de una clasificación estándar de valores para cada índice (EWS, Emergency Warning Score System), habilitando la toma de decisiones tales como su reemplazo. Para finalizar con este objetivo se pretende crear un servicio de suscripción a eventos a través de plataformas que realizan las funciones de puente de comunicación entre la red de blockchain y las entidades que necesiten tratar con los datos relacionados con los eventos emitidos por los smart contracts.

IV. CONCLUSIONES

Los dos grandes retos que debe afrontar la IoMT son la seguridad y la eficiencia energética. Ambos se corresponden con los dos grandes objetivos perseguidos en este trabajo de investigación en curso. Tras haber descrito brevemente los trabajos más recientes relacionados con la temática de la presente investigación, se han especificado los avances ya alcanzados por nuestro equipo y se han descrito aquellas líneas en las que actualmente se está trabajando, así como los hitos que se pretenden alcanzar.

AGRADECIMIENTOS

Este proyecto (RTI2018-097763-B-I00) está parcialmente financiado por: FEDER/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación. Este proyecto ha sido parcialmente financiado por la Agencia Estatal de Investigación (AEI). PID2020-115323RB-C32/ AEI / 10.13039/501100011033.

REFERENCIAS

[1] M. Asam, T. Jamal, M. Adeel, A. Hassan, S. Aziz Butt, A. Ajaz and M. Gulzar, “Challenges in Wireless Body Area Network”, International Journal of Advanced Computer Science and Applications(IJACSA), 10(11), 2019.

[2] Liu, Q., Mkongwa, K.G. and Zhang, C. “Performance issues in wireless body area networks for the healthcare application: a survey and future prospects”. SN Appl. Sci. 3, 155 (2021).

[3] R. Matsuo, T. Nabetani, H. Tanakay, W. H. Chin, and S. Subramani, “Performance of simple and Smart PHY/MAC mechanisms for Body Area Networks,” in Proc. IEEE ICC, London, UK, 2015, pp. 501–506.

[4] L. Ruan, M. Dias, and E. Wong, “SmartBAN with Periodic Monitoring Traffic: A Performance Study on Low Delay and High Energy Efficiency,” IEEE J. Biomed. Health Inform., vol. 22, no. 2, pp. 471–482, March 2018.

[5] Z. Liu, B. Liu, Ch. Chen, and Ch. W. Chen, “Energy-Efficient Resource Allocation with QoS Support in Wireless Body Area Networks,” in Proc. IEEE Globecom, San Diego, CA, USA, 2015, pp. 1–6.

[6] B. Liu, Z. Yan, and Ch. W. Chen, “Medium Access Control for Wireless Body Area Networks with QoS Provisioning and Energy Efficient Design,”IEEE Trans. Mobile Comput., vol. 16, no. 2, pp. 422–434, Feb. 2017.

[7] F. Solt et al., “Energy Efficient Heartbeat-Based MAC Protocol for WBAN Employing Body Coupled Communication,” in IEEE Access, vol. 8, pp. 182966-182983, 2020.

[8] R. Khan, M. M. Alam and M. Guizani, “A Flexible Enhanced Throughput and Reduced Overhead (FETRO) MAC Protocol for ETSI SmartBAN,” in IEEE Transactions on Mobile Computing.

[9] S. Kosunalp, “SMAC Protocols for Energy Harvesting Wireless Sensor Networks: Survey”, in ETRI journal,Vol. 37, Is. 4, pp. 804-812,August 2015, Wiley

[10] Hao Y, Peng L, Lu H, Hassan MM, Alamri A. Energy Harvesting Based Body Area Networks for Smart Health. Sensors (Basel). 2017;17(7):1602. Published 2017 Jul 10.

[11] C. Habib, A. Makhoul, R. Darazi and C. Salim, “Self-Adaptive Data Collection and Fusion for Health Monitoring Based on Body Sensor Networks,” in IEEE Transactions on Industrial Informatics, vol. 12, no. 6, pp. 2342-2352, Dec. 2016.

[12] G. Prisco, “The blockchain for healthcare: Gem launches gem health network with philips blockchain lab,” Bitcoin Magazine, 2016.

[13] K. Hasan, K. Biswas, K. Entenam, U. Ahmed, Md. S. Islam, “Challenges of Integrating blockchain in Wireless Body Area Network”, Noviembre 2018.

[14] Griggs, K.N., Ossipova, O., Kohlios, C.P. et al. “Healthcare blockchain System Using Smart Contracts for Secure Automated Remote Patient Monitoring”, J. Med. Sys.t 42, 130 (2018).

[15] H. Syeda, Z. Kazmi, F. Nazeer, S. Mubarak, S. Hameed, A. Basharat, N. Javaid “Trusted Remote Patient Monitoring using blockchain-based Smart Contracts”, Advances on Broad-Band Wireless Computing, Communication and Applications. BWCCA 2019. Lecture Notes in Networks and Systems, vol 97. Springer, Cham.

[16] P. P. Ray, D. Dash, K. Salah and N. Kumar “Blockchain for IoT-Based Healthcare: Background, Consensus, Platforms, and Use Cases”, in IEEE Systems Journal.

[17] R. Kumari, P. Nand and R. Astya, “Integration of blockchain in WBAN,” 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 144-149.

[18] Yaqoob, I., Salah, K., Jayaraman, R. et al. “Blockchain for healthcare data management: opportunities, challenges, and future recommendations”. Neural Comput and Applic (2021).

[19] R. Wasim Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, S. Ellahham, M. Omar, “The role of blockchain technology in telehealth and telemedicine”, International Journal of Medical Informatics, Volume 148, 2021, 104399, ISSN 1386-5056,

[20] J. Ramis-Bibiloni and L. Carrasco-Martorell, “An energy-efficient and delay-constrained resource allocation scheme for periodical monitoring traffic in SmartBANs,” in Proc. IEEE BioCAS, Turin, Italy, 2017, pp.1–4.

[21] J. Ramis-Bibiloni and L. Carrasco-Martorell, “Energy-Efficient and QoS-Aware Link Adaptation With Resource Allocation for Periodical Monitoring Traffic in SmartBANs,” in IEEE Access, vol. 8, pp. 13476-13488, 2020.

[22] J. Ramis-Bibiloni and L. Carrasco-Martorell, “Energy Harvesting Effect on the Sensors Battery Lifespan of an Energy Efficient SmartBAN Network,” IWCMC, Harbin, China, 2021, pp.1–6.

[23] M. Mut-Puigserver, M. Payeras-Capellà and M. Cabot-Nadal, ”Blockchain-Based Fair Certified Notifications”, Data Privacy Management, Cryptocurrencies and blockchain Technology”, LNCS 11025, Springer International Publishing, pp 20-37, 2018.

[24] M. Mut-Puigserver, M. Payeras-Capellà, M. Cabot-Nadal, ”Blockchain-based Contract Signing Protocol for Confidential Contracts,” Proceedings of the IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), IEEE eXpress Conference Publishing, ISBN: 978-1-7281-5052-9, ISSN: 2161-5330, November 2019.

[25] M. Mut-Puigserver, M. A. Cabot-Nadal and M. Payeras- Capellà,”Removing the Trusted Third Party in a Con- fidential Multiparty Registered eDelivery Protocol using blockchain,” in IEEE Access, 2020.

[26] M. Payeras-Capellà, M. Mut-Puigserver and M. A. Cabot-Nadal, Blockchain-Based System for Multiparty Electronic Registered Delivery Services, in IEEE Access, vol. 7, pp. 95825-95843, 2019.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Arquitectura LoRaWAN para entornos sin cobertura

Félix Delgado-Ferro, Jorge Navarro-Ortiz, Natalia Chinchilla-Romero, Juan José Ramos-Muñoz
Departamento de Teoría de la Señal, Telemática y Comunicaciones,
Universidad de Granada
C/ Periodista Daniel Saucedo Aranda, s/n. ETSI Informática y de Telecomunicación
felixdelgado@correo.ugr.es, jorgenavarro@ugr.es, nataliachr@ugr.es, jjramos@ugr.es

Este trabajo describe el diseño, desarrollo e implementación de una red pensada para emplearse en zonas sin cobertura como montaña o desierto mediante tecnologías de comunicaciones inalámbricas (*Bluetooth Low Energy* - BLE y *Low Power Wide Area Network* -LPWAN). En la implementación se emplean elementos de bajo consumo para la creación de redes de amplio rango de cobertura.

***Palabras Clave*- Internet of Things, LoRaWAN, Bluetooth Low Energy, Android**

I. INTRODUCCIÓN

Las comunicaciones inalámbricas son a día de hoy el tipo de tecnología más empleada en la sociedad. Por este motivo, éstas siguen prosperando y alcanzando límites que años atrás ni imaginábamos. Estos avances nos permiten la comunicación tanto con otras personas como con las máquinas y el entorno.

Hoy en días, existen aún zonas sin cobertura en España, normalmente zonas rurales donde no llegan las operadoras o zonas que se conocen como sombras. Debido a esta problemática y motivado por la participación de grupos de emergencias como Protección Civil en zonas sin cobertura ni conectividad (e.g. Sierra Morena), nos planteamos la posibilidad de crear una red de interconexión móvil para dar soporte a estos equipos de emergencias.

El proyecto se centra en el uso de diversas tecnologías de comunicación inalámbrica como Bluetooth y LoRaWAN para garantizar la conectividad en zonas sin cobertura. Estas tecnologías se seleccionaron debido a que emplean bandas de frecuencias sin licencias (ISM) y presentan características como envío de datos con bajo consumo energético y rentabilidad.

II. ESTADO DEL ARTE

Actualmente, existen algunas soluciones comerciales alternativas para la conectividad en zonas sin cobertura como SPOT X [1] que permite la conectividad empleando un hardware específico al que el móvil se conecta mediante Bluetooth, Uepaa! App [2] que permite envío de mensajes de emergencias utilizando *multihop networks* y Beartooth [3] que permite crear redes amplias empleando hardware específico para el envío de localización y mensajes entre usuarios.

III. FUNDAMENTOS TEÓRICOS

A. *Bluetooth Low Energy*

Bluetooth Low Energy (BLE) es parte de la versión 4.0 del núcleo de especificaciones Bluetooth [4]. Fue diseñado con el objetivo de optimizar costes con bajo ancho de banda, baja potencia y baja complejidad.

Existen diferentes versiones de dispositivos dependiendo del tipo de conexiones que soporten [5]. Estas son Bluetooth clásico (solo BR/EDR), monomodo (solo BLE) y dual (ambos).

El protocolo Bluetooth Low Energy (BLE) mantiene una estructura de tres partes: controlador, host y aplicación. Estos albergan todas las funcionalidades de su pila de protocolos.

- Controlador: gestiona las comunicaciones entre dispositivos BLE mediante maestro-esclavo.
- Host: gestiona las claves de seguridad, el intercambio de datos entre aplicaciones y controla las conexiones.
- Aplicación: proporciona un servicio al usuario.

B. LoRaWAN

LoRaWAN es una especificación que define tanto el protocolo de comunicaciones como la arquitectura de red. El protocolo LoRaWAN [6] describe las siguientes capas:

- Capa física: establece la comunicación y opera en la banda ISM (868 MHz en Europa). Además, utiliza la modulación LoRa para establecer enlaces de comunicación a largo alcance con poca potencia.
  Esto es posible al emplear el parámetro *spreading factor* (SF) que permite buscar un compromiso entre robustez y velocidad, modificando además el número de canales ortogonales donde transmitir sin colisiones.
- Capa MAC: gestiona el acceso al medio, es decir, los canales y parámetros de conexión. En la especificación, existen tres tipos de dispositivos dependiendo de la bidireccionalidad y los tiempos de espera [6] (Clase A, B y C).

Por otro lado, LoRaWAN define una arquitectura de red y las funcionalidades de seguridad que implementa. La red está formada por cuatro tipos de dispositivos [6]:

- Mota: dispositivos integrados de comunicación de baja potencia.
- Gateway: pasarela de las transmisiones entre motas y servidores de red.
- Servidor de Red: autentica los nodos, comprueba la integridad de los mensajes y los enruta a los servidores de aplicación.
- Servidor de Aplicación: incorpora mecanismos de confidencialidad de mensajes y gestiona las solicitudes.

Además, permite la activación de los dispositivos mediante activación por personalización (ABP), en la que las claves están pregrabadas en la mota, o mediante activación por aire (OTAA), en la que las claves se generan tras el envío de señalización [6].

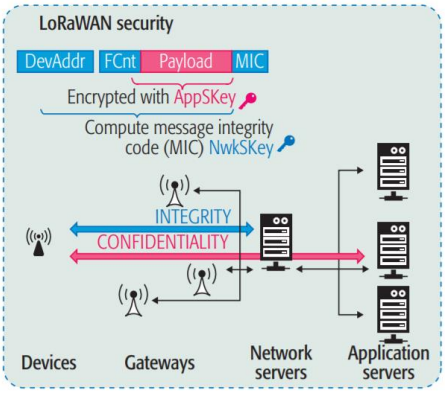


Fig. 1. Arquitectura de la red LoRaWAN [7]

IV. Diseño de Red

La idea fundamental de la red es que permita transmitir información (mensajes y localización) desde cualquier dispositivo móvil al nodo central, incluso si estos se encuentran en zonas sin cobertura.

La red TeamUp se implementa fundamentalmente sobre una arquitectura LoRaWAN. Esta arquitectura será el núcleo de nuestra red e integramos nuevas funcionalidades y elementos a la red como una aplicación móvil, un servidor propio en el nodo central y una interfaz web de gestión de la red.

Respecto a la conectividad entre los dispositivos que conforman la red y las funcionalidades que cada uno atribuye se especifican a continuación.

- Smartphone: integra la aplicación *TeamUp*.
- Mota: permite el envío/recepción de mensajes hacia/desde el móvil usando comunicaciones BLE y cambia el mensaje a LoRa para la retransmisión hasta el gateway o viceversa.
- Gateway: se encarga de la recepción de los mensajes LoRa de las motas y lo retransmite al nodo central empleando Wi-Fi o viceversa.
- Nodo central: permite trabajar en modo estático o dinámico, siempre que ofrezca cobertura al equipo de emergencias. Se encarga de actualizar los servidores (chirpstack & propietario). Estos servidores se emplean para el almacenamiento de los mensajes que se enviaron desde la aplicación móvil, permiten peticiones web para gestión e incluyen funcionalidades de reenvío a otros usuarios.
- Interfaz Web: facilita la visualización de los datos a través de un navegador web e incluye funcionalidades como registro de usuario, visualización de mensajes o localización, etcétera.

La red debe permitir el envío de mensajes entre miembros del equipo de emergencias y la visualización para gestión de la red vía web.

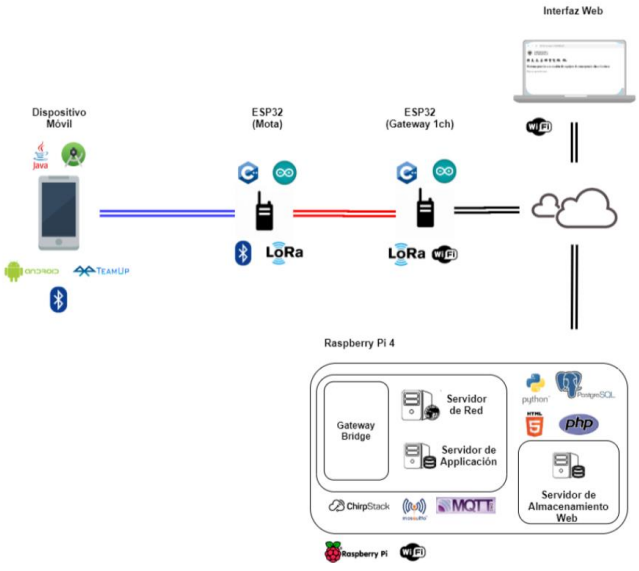


Fig. 2. Diseño de la red TeamUp [8]

V. Implementación de la Red

A. App TeamUp

La aplicación TeamUp se ha diseñado de forma que visualmente y funcionalmente sea intuitiva para los usuarios y sea ameno el navegar entre las seis pantallas.

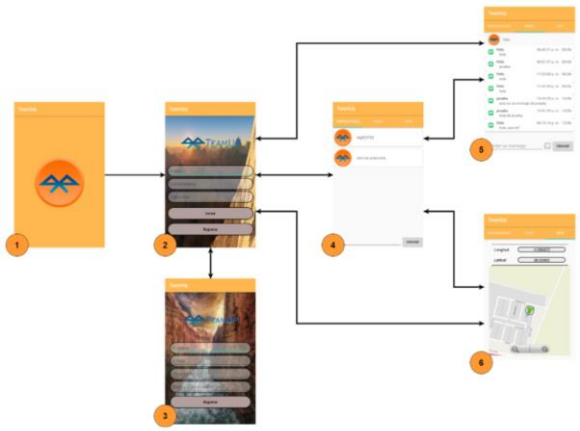


Fig. 3. App TeamUp [8]

Estas pantallas permiten diferentes funcionalidades como registro, login, envío y recepción de mensajes y conectarse a las motas vía BLE. La implementación de estas funcionalidades se ha realizado modularizando dependiendo de la funcionalidad y empleando Android Studio de forma que fuese compatible con el 99% de los dispositivos Android.

B. Motas

Las motas se han implementado empleando diferentes tipos de dispositivos debido a las características que incorporaban cada uno de ellos.

Primero, empleamos ESP32-LoRa que es un dispositivo de clase A y nos permite reducir costes económicos y energéticos, mientras que la cobertura es de gran alcance, pero con baja capacidad de transmisión. Por otro lado, se ha empleado Pycom FiPy que son dispositivos clase C, es decir, son más costosos económicamente y energéticamente, pero permanece escuchando en todo momento y mejora el tiempo de interacción y la latencia.

C. Gateway

El gateway se encarga del envío y recepción de los mensajes LoRA con las motas y la retransmisión mediante Wi-Fi a los servidores. En el despliegue se han empleado tres tipos de dispositivos que ofrecen distintas variedades en la implementación.

- IMST Gateway Lite: realiza un barrido de los 8 canales disponibles, obteniendo altas capacidades. Utiliza una Raspberry Pi 1.
- ESP32 - Single Channel Gateway: es una implementación más económica, aunque reduce el rendimiento dado que es un dispositivo de clase A. La implementación se ha realizado empleando la librería desarrollada por Marteen Westenberg que encontramos en Github [9].
- LoPy 4.0 - Nano Gateway: implementa la misma funcionalidad que el Single Channel Gateway. La implementación es similar, pero empleando librerías de Python [10].

D. Servidores Chirpstack

La plataforma Chirpstack es el núcleo de la red LoRaWAN, implementando los servidores de red y aplicación (además de otras entidades relacionadas) y, por tanto, de la red TeamUp. Estos servidores se encargan de la recepción, desencriptado y almacenamiento de los mensajes. Además, implementan mecanismos de seguridad basados en confidencialidad e integridad a dos niveles, es decir, los mensajes mantienen la integridad hasta los servidores de red y la confidencialidad hasta los servidores de aplicación.

La instalación y configuración de los servidores Chirpstack consiste en la instalación de las dependencias y los paquetes software de Chirpstack, la configuración de la organización, perfil de servicio, aplicación, perfil de dispositivo y, finalmente, configurar los dispositivos.

E. Servidor Propietario

El servidor propietario tiene dos funciones que podemos distinguir fácilmente. La primera se encarga de la actualización de la base de datos y reenvío en caso de ser mensajes para el grupo de emergencias. La segunda función consiste en el despliegue de un servicio web para la visualización y gestión de la red por parte del administrador de la misma. Este servidor se ha implementado empleando Python y lenguajes de peticiones web como PHP. Además, este servidor interactúa con la plataforma Chirpstack a través del protocolo MQTT y de la REST API.

VI. Pruebas de Concepto

Las pruebas de concepto se han realizado para comprobar las capacidades de cada uno de las conexiones y la cobertura que podía ofrecer la red. Una vez la red estaba completamente integrada, se han realizado comprobaciones de las funcionalidades. A modo de ejemplo completo, se va a comentar una prueba que consiste en el inicio de sesión de un usuario en la red *TeamUp* [8]. Esta prueba sigue los siguientes pasos:

1. Puesta en marcha del servidor: se inicia el servidor propietario con el comando *python* para que se actualice la base de datos privada constantemente.
2. Conexión LoRaWAN: se espera a que la mota y el gateway establezcan la conexión automáticamente.
3. Conexión BLE: registramos al usuario en la aplicación y se escanean los dispositivos BLE desde. Posteriormente, se realiza la conexión con la mota. En este momento, se envía automáticamente el mensaje de *Login* para iniciar sesión y guardar el registro en la base de datos. Este mensaje tiene un formato específico: #L# [usuario] [contraseña]
4. Recepción del mensaje *Login ACK*: el mensaje llega al servidor y es procesado. Al detectar que se trata de

un mensaje de *Login* se contesta al usuario mediante un mensaje *Login ACK* que tiene un formato #LACK.

5. Mensajes en la Interfaz Web: el administrador de la red tiene la posibilidad de visualizar los mensajes y entre ellos, puede ver los mensajes de inicio de sesión con los formatos especificados.

Registro de los mensajes recibidos por el servidor de aplicación LoRaWAN:

id	dev_eui	message	time	direction
1	70b3d54994de968f	#L# Felix 123456	2020-07-03 22:26:36.542257+01	uplink
2	70b3d54994de968f	#LACK#	2020-07-03 22:26:37.339966+01	downlink
3	70b3d54994de968f	Esto es una prueba	2020-07-03 22:27:30.233449+01	downlink

Fig. 4. Mensajes en el Interfaz Web [8]

VII. CONCLUSIONES

Este trabajo presentaba una motivación personal y una alta complejidad debida a la interconexión de miembros de un equipo de emergencias en zonas sin cobertura.

Principalmente, se hizo un estudio de las tecnologías disponibles y se valoraron las opciones que tenían estas en el proyecto, llegando a elegir LoRaWAN por el gran rango de cobertura que ofrece y su bajo consumo energético y Bluetooth Low Energy para la conexión del smartphone dado que éstos no soportan LoRaWAN.

Posteriormente, se diseñó una red que permitiera suplir los problemas de conectividad. La implementación de la red se centra en un núcleo de red LoRaWAN (Chirpstack) al que se han incluido funcionalidades extras como una aplicación móvil, un servidor propietario para gestionar mensajes y ofrecer un servicio web e una interfaz web.

Finalmente, se realizaron un conjunto de pruebas para comprobar el funcionamiento de la red. Ésta funciona como se esperaba en zonas sin cobertura, permitiendo a los usuarios enviar y recibir mensajes a través de la red. Por tanto, afirmamos que la red es 100% operativa, aunque este es el primer paso.

VIII. TRABAJOS FUTUROS

La red está en pleno funcionamiento y ese es el punto perfecto para realizar pruebas y análisis de rendimientos. Las pruebas de rendimiento podemos dividirlas en dos experimentos dependiendo de la funcionalidad del mismo.

1. Comprobación del rendimiento y rango de cobertura en distintos escenarios, es decir, se pretende realizar un análisis de rendimiento en función de la distancia para escenarios naturales, rurales y urbanos. Estos análisis se realizarán en las localizaciones marcadas sobre la figura 5.
2. Caracterización del rendimiento dependiendo de la movilidad de la red, es decir, la red es centralizada, pero permite que el nodo central sea dinámico y se desplace. Por tanto, es interesante comprobar como el movimiento tanto de motas como del nodo central a la vez afectan al rendimiento de la red. Obviamente se compararán los resultados con las mediciones en estático.

Aparte de estos análisis que se pretenden realizar, existen posibles mejoras sobre la red o implementaciones alternativas que se mencionan a continuación:

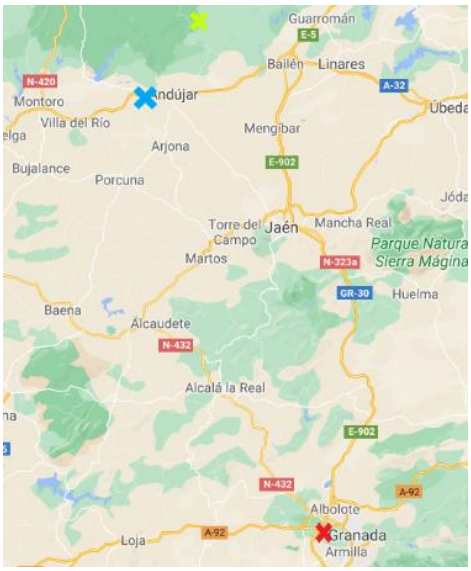


Fig. 5. Localización de las pruebas de rendimiento

- Mejora el diseño de la interfaz web dado que solo se ha llegado a implementar la parte funcional, pero no es atractivo para el usuario final.
- Implementación y comprobación de la ruta de un usuario desde la interfaz web empleando Google Maps, es decir, hacer un *tracking* de los usuarios de la red en tiempo real.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por la Agencia Andaluza del Conocimiento (proyecto A-TIC-241-UGR18) y el Ministerio de Economía y Competitividad (proyecto TEC2016-76795-C6-4-R).

REFERENCIAS

[1] SPOT X Globalstar. Disponible: https://www.globalstar.com/es-la/products/personnel-safety/spotx (visitado 14-04-2020)

[2] Uepaa Safety App Losung. 'Die clevere App-Losung fur Alleinarbeiter mit Totmannfunktion. Leichtgewichtige Smartphone App mit 24/7 Notrufzentrale". url: https://safety.uepaa.ch/de/ (visitado 14-04-2020)

[3] Official Page of Beartooth. url: https://beartooth.com (visitado 14-04-2020)

[4] Bluetooth Technology. 'Core Specifications'. Disponible: https://www.bluetooth.com/specifications/bluetooth-core-specification/ (visitado 14-04-2020)

[5] Bluetooth Technology. 'Radio Versions'. Disponible: https://www.bluetooth.com/learn-about-bluetooth/ bluetoothtechnology/ radio-versions/ (visitado 14-04-2020)

[6] LoRaWAN® Specification v1.1, LoRa Alliance®. Disponible: https://lora-alliance.org/resource_hub/lorawan-specification-v1-1/ (visitado 14-04-2020)

[7] Navarro-Ortiz, Jorge and Sendra, Sandra and Ameigeiras, Pablo and Lopez-Soler, Juan M. (2018,02). 'Integration of LoRaWAN and 4G/5G for the Industrial Internet of Things'. IEEE Communications Magazine. volumen 56, págs. 60-67. Disponible: https://ieeexplore.ieee.org/document/8291115 (25-06-2021)

[8] Delgado Ferro, Félix. 'Interconexión de Miembros de un Equipo de Emergencias en Entornos sin Cobertura usando Dispositivos Móviles'. Disponible: https://wpd.ugr.es/ jorgenavarro/ thesis/2020 TFG FelixDelgadoFerro.pdf (visitado 25-06-2021)

[9] Version 5 of Single Channel LoRa Gateway, Jac Kersing GitHub Disponible: https://github.com/kersing/ESP-1ch-Gateway-v5.0 (visitado 20/05/2020)

[10] Pycom. 'API Reference for Nano Gateway'. Disponible: https://docs.pycom.io/tutorials/lora/lorawan-nano-gateway/ (visitado 14-04-2020)


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Implementación en el Espacio de Usuario del Protocolo de Encaminamiento AODVv2

Sergio Machado, Israel Martín-Escalona, Enrica Zola, Francisco Barceló-Arroyo, Javier Ozón
Departamento de Ingeniería Telemática,
Universidad Politècnica de Catalunya, UPC BarcelonaTECH
Barcelona, Spain.
{sergio.machado, israel.martin, enrica.zola, francisco.barcelo, francisco.javier.ozon}@upc.edu

El protocolo AODV es un protocolo reactivo de encaminamiento para redes MANET que se utiliza frecuentemente como referencia tanto para desarrollar nuevos protocolos de encaminamiento en redes ad hoc como para evaluar su rendimiento. Aunque el protocolo está presente en varios simuladores de red (por ejemplo, ns2, OMNeT++, etc.), existen pocas implementaciones que puedan emplearse en condiciones reales para la investigación de campo o de evaluación. Este documento presenta una implementación en el espacio de usuario de la última versión del protocolo, el AODVv2, que se puede utilizar en cualquier dispositivo capaz de ejecutar Linux. El objetivo del proyecto ha sido el desarrollo de una implementación de AODVv2 de código abierto y fácil mantenimiento, para ser utilizada con fines experimentales por la comunidad científica. El presente documento proporciona una descripción de los principales criterios de diseño y codificación considerados para implementar el protocolo, y expone además las principales pruebas que se han realizado con el fin de verificar su correcto funcionamiento.

***Palabras Clave*—AODVv2, MANET, ad hoc, protocolos de encaminamiento, implementación, Linux, espacio de usuario.**

I. INTRODUCCIÓN

Las Mobile Ad Hoc Networks (MANETs) son redes inalámbricas carentes de infraestructura en las que los nodos móviles se comunican entre sí de forma descentralizada. En una red MANET los nodos desempeñan distintos papeles, puesto que pueden ser tanto emisores como receptores de un determinado flujo de información, así como intermediarios en la comunicación entre otros pares de nodos. Un protocolo de encaminamiento debe determinar en cada momento qué nodos se ocupan de la retransimisión de un determinado flujo de información, i.e. de reenviar una trama al siguiente nodo de la ruta que une el emisor con el receptor.

This work was supported by the Spanish Government and ERDF through CICYT project PGC2018-099945-B-I00.

Los protocolos de encaminamiento de las MANETs pueden clasificarse conforme distintos criterios. Una de las clasificaciones más comunes divide los protocolos en encaminamiento no jerárquico y encaminamiento jerárquico. En el primer caso, las reglas de encaminamiento se aplican uniformemente a todos los nodos de la red. En el segundo, los nodos de la red se dividen en distintos grupos y los protocolos aplican distintos criterios conforme el encaminamiento comunica nodos del mismo grupo o bien nodos de grupos distintos. A su vez, los protocolos de encaminamiento no jerárquico se clasifican en tres categorías: proactivos, reactivos e híbridos. Los protocolos proactivos monitorizan regularmente la topología de red y aplican soluciones clásicas de encaminamiento en redes fijas, tanto de tipo vector distancia como de estado del enlace. Mientras los protocolos proactivos disponen en cada momento de un encaminamiento posible entre cualquier par de nodos, los protocolos reactivos funcionan bajo demanda, es decir, activan un proceso de encaminamiento cada vez que un paquete se ha de enviar desde un nodo fuente a un nodo destino. El encaminamiento conserva su validez durante un determinado periodo de tiempo y se emplea hasta que expira dicho plazo o el protocolo comprueba que ha dejado de funcionar. En este caso, el protocolo determinará un nuevo encaminamiento entre el nodo emisor y el receptor. Dado que ambos mecanismos presentan ventajas y desventajas, los protocolos híbridos tratan de combinar los beneficios de ambas aproximaciones.

Con independencia de la estrategia seguida, i.e. jerárquica o no jerárquica, los protocolos de encaminamiento para MANETs pueden clasificarse a su vez conforme dispongan o no de la posibilidad de obtener y, por tanto, aprovechar la localización de los nodos en la estrategia de encaminamiento. Esta política de localización de los nodos se ha demostrado además imprescindible para un correcto escalado de dichos protocolos [3].

Ad hoc On Demand Distance Vector Routing (AODV) [4] es un protocolo de encaminamiento MANET

que establece y mantiene rutas entre nodos mediante mensajes de petición de ruta (route request) y mensajes de respuesta de ruta (route reply). Cuando se ha de establecer una ruta entre dos nodos, el nodo emisor inunda los nodos vecinos con un mensaje de petición de ruta; cada uno de los nodos vecinos retransmite, a su vez, el mensaje de petición a sus vecinos y así sucesivamente hasta alcanzar el nodo destino. En este punto, el nodo destino responde con un mensaje de respuesta de ruta que recorre en sentido inverso el camino previamente trazado por el mensaje de petición de ruta.

Aunque el protocolo AODV version 2 (AODVv2) [5] introduce algunas mejoras sobre AODV, el mecanismo de encaminamiento es fundamentalmente el mismo en ambos casos. Algunas de las mejoras más relevantes contempladas por AODVv2 son: inundación optimizada de los mensajes de petición de ruta con el fin de reducir la sobrecarga (overhead) debida al encaminamiento; ampliación del número de métricas que pueden considerarse; empleo de una estructura Type-Length-Value (TLV) para la codificación de los campos de datos, que añade flexibilidad a la información que ha de enviarse, tal y como define el RFC Generalized MANET Packet/Message Format [6]; secuenciación de los mensajes Route Request (RREQ) y Route Reply (RREP); mecanismo que evita que los nodos intermedios respondan a los mensajes RREQ cuando conocen una ruta al destino solicitado; y uso de mensajes Route Error (RERR) para avisar de la caída de un enlace así como de mensajes Route Reply Acknowledgment (RREP_ACK) para comprobar el funcionamiento de los enlaces bidireccionales de la red.

AODVv2 es un protocolo que se toma con frecuencia como referencia en la evaluación del rendimiento de nuevos protocolos de encaminamiento. Además, los investigadores suelen tomar AODVv2 como punto de partida para desarrollar sus propios protocolos [7]. Actualmente se encuentran disponibles en la literatura varias implementaciones del algoritmo; sin embargo, la mayoría de dichas implementaciones están concebidas únicamente con fines de simulación [8] [9], han quedado obsoletas [10] [11] o bien se restringen a ciertas configuraciones antiguas de software. En [12], los autores presentan una implementación de AODV en Python, mientras que en [13] se implementa AODV mediante Exata Emulator y Matlab. Con respecto a las implementaciones de AODVv2, los autores de [9] detallan la implementación del protocolo en el programa de simulación OMNeT++ dentro del marco INET. Sin embargo, dado que el protocolo AODVv2 ha evolucionado desde 2008, se ha hecho necesaria una actualización para garantizar la fiabilidad de futuras pruebas. Más recientemente, los autores en [14] presentan una adaptación de la implementación de AODV-UU [10] a las versiones más recientes del kernel. En el presente trabajo, hemos adoptado el enfoque de [14], que usa una implementación del protocolo en el espacio de usuario para ganar independencia del kernel. Nuestro objetivo es desarrollar una nueva implementación, de código abierto y fácil de mantener, del protocolo AODVv2 dirigido a todos los investigadores que trabajan en protocolos de encaminamiento MANET.

El objetivo de este artículo es proporcionar una descripción completa de esta nueva implementación del protocolo AODVv2. Este trabajo está dividido de la siguiente manera. En la Sección II se describen los mensajes y el funcionamiento del protocolo AODVv2. La sección III presenta los criterios de diseño de nuestra implementación de AODVv2 en el espacio de usuario de Linux. La sección IV describe las pruebas y los casos que se han planteado con el fin de comprobar el correcto funcionamiento de la implementación. Finalmente, en la Sección V se exponen las principales conclusiones del proyecto y se plantean posibles mejoras futuras.

II. DESCRIPCIÓN GENERAL DEL PROTOCOLO

El protocolo AODVv2 se ocupa de la gestión dinámica de una red inalámbrica ad hoc con el objeto de garantizar, cuando es físicamente posible, la disponibilidad de una ruta entre cualquier par de dispositivos conectados a la red. AODVv2 sustituye al protocolo AODV e incluye algunas mejoras, como una mayor fiabilidad de las pruebas de los enlaces inalámbricos, la transmisión y procesado de respuestas a las peticiones de ruta y un mayor rango de búsqueda de rutas. Como su predecesor, AODVv2 es un protocolo de encaminamiento de vector distancia, en el que los datos de ruta incluyen el siguiente nodo al que se debe reenviar un paquete y el coste (métrica) de dicho salto. De forma predeterminada, la métrica considera el número de saltos de la ruta. De este modo, cuando existe más de un camino entre el emisor y el receptor, el protocolo escoge aquel que contiene un menor número de enlaces.

En Fig. 1 y Fig. 2 se describe la operativa básica del protocolo. La Fig. 1 muestra un escenario donde el nodo emisor *S* pretende enviar un mensaje al nodo destino *D*, que no está conectado directamente a *S*. En la figura, las líneas discontinuas representan transmisiones broadcast. AODVv2 es un protocolo reactivo. De este modo, cuando un nodo emisor no dispone de una ruta al nodo destino, envía un mensaje broadcast RREQ con el objeto de descubrir una ruta que comunique el origen con el destino. El mensaje RREQ será recibido, procesado y reenviado por los nodos intermedios (llamados también nodos retransmisores) a lo largo del camino entre el nodo emisor y el receptor. En Fig. 1 el mensaje broadcast RREQ enviado por *S* es recibido por los nodos *A* y *E*. Dado que ninguno de los dos es el nodo destino del mensaje RREQ, *A* y *E* volverán a enviar el mensaje RREQ a sus nodos vecinos. Si un nodo recibe el mismo RREQ en más de una ocasión, solo lo reenviará la primera vez y lo descartará en las ocasiones sucesivas. El mensaje RREQ incluye un número de secuencia que permite identificar los mensajes RREQ duplicados.

En Fig. 1, el nodo *A* descarta el mensaje RREQ reenviado por el nodo *E*, puesto que se habrá identificado como mensaje duplicado gracias al número de secuencia. Cuando un nodo recibe un mensaje RREQ útil, i.e. que no debe ser descartado, el nodo almacena la ruta inversa que

lo comunica con el nodo que ha emitido inicialmente el RREQ, en este caso el nodo *S*. En Fig. 1, el nodo *A* sabrá que el nodo *S* es uno de sus nodos adyacentes, si es que no lo sabía previamente, y que por tanto puede alcanzar dicho nodo mediante un único salto. De modo análogo, el nodo *B* reconocerá que puede llegar al nodo *S* a través del nodo *A*. Asimismo, el nodo *C* descubrirá que puede llegar a *S* a través del nodo *B*. Finalmente, el mensaje RREQ llegará al nodo destino *D*, que sabrá que puede comunicarse con el nodo *S* a través del nodo *C*. Todos los nodos implicados almacenan localmente la correspondiente información en sus respectivas tablas de encaminamiento.

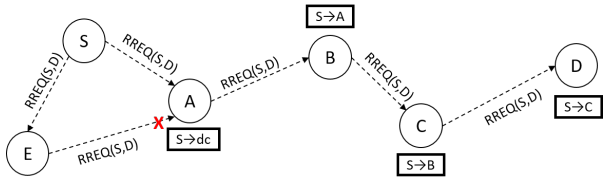


Fig. 1. Ejemplo de inundación RREQ.

Fig. 2 muestra el mensaje RREP enviado por el nodo destino con el fin de completar el proceso de creación de la ruta. Tras recibir el mensaje RREQ creado por el nodo *S*, el nodo *D* responde con un mensaje RREP. A diferencia de lo que sucede con RREQ, los mensajes RREP no se envían en modo broadcast sino en modo unicast, representado en la figura mediante flechas continuas. Esto es así puesto que los nodos intermedios han almacenado localmente el camino de vuelta al nodo emisor conforme el mensaje RREQ recorría el camino de *S* a *D*. De este modo, el nodo receptor *D* enviará el mensaje RREP únicamente al nodo *C*, el cual, a su vez, lo reenviará únicamente al nodo *B* y así sucesivamente hasta que el mensaje RREP llegue al nodo emisor *S* que comenzó todo el proceso. Finalmente, una vez el nodo *S* haya recibido el mensaje RREP, se considerará que la ruta ha sido correctamente establecida y los paquetes almacenados en la cola del emisor *S* empezarán a ser enviados a su destino *D*.

Complementariamente, AODVv2 permite comprobar el funcionamiento de los enlaces cuando crea una ruta. Para ello, cuando un nodo envía o retransmite un mensaje RREP, debe enviar también un mensaje RREP_ACK al siguiente nodo, quien a su vez debe responderle con un mensaje RREP_ACK. Este mecanismo, que debe omitirse cuando se alcanza el límite del ancho de banda del enlace reservado para mensajes de control, aparece en Fig. 2 en el enlace que une los nodos *D* y *C* y se obvia en el resto por simplicidad.

AODVv2 emplea asimismo mensajes RERR para informar sobre la caída de enlaces en la red. Existen tres casos en los que el protocolo transmite un mensaje RERR. En primer lugar, cuando un nodo recibe un paquete que ha de retransmitir pero no dispone de una ruta válida a la dirección de destino. En este caso, el nodo envía un mensaje unicast RERR a la dirección origen del paquete. En segundo lugar, cuando un nodo recibe un mensaje RREP pero no dispone de una ruta al nodo que creó

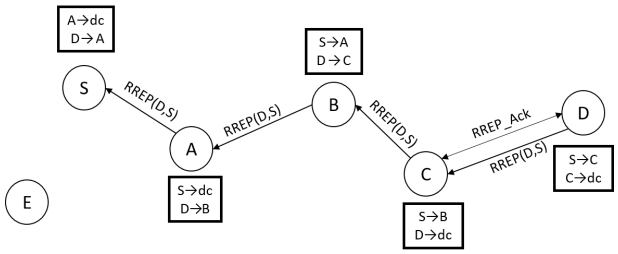


Fig. 2. Ejemplo de RREP.

el correspondiente mensaje RREQ, i.e. una ruta al nodo emisor. En este caso, el nodo envía un mensaje unicast RERR al nodo que creó el mensaje RREP, i.e. al nodo destino de la comunicación. Nótese que tanto el paquete de información en el primer caso como el mensaje RREP en el segundo, son descartados por el nodo intermedio que crea el mensaje RERR. El tercer supuesto en que se transmite un mensaje RERR ocurre cuando un nodo detecta la caída de un enlace que puede desactivar distintas rutas entre nodos. En este caso, el nodo envía un mensaje multicast RERR a todos sus nodos adyacentes.

Con el fin de evitar congestiones, el tráfico de control de AODVv2 solo puede consumir el 10% de la capacidad de cada enlace. Cuando supera este umbral, el protocolo prioriza la detección y notificación de la caída de enlaces y también de la presencia de rutas incorrectas, con el objeto de encontrar rápidamente nuevas rutas disponibles para la transmisión de datos.

III. IMPLEMENTACIÓN DEL PROTOCOLO AODVV2

Esta sección presenta las decisiones de diseño más relevantes en la implementación del protocolo AODVv2. Dicha implementación ha sido concebida para ser ejecutada en dispositivos con el sistema operativo Linux. Hay dos aproximaciones para implementar un protocolo de encaminamiento para Linux según el tipo de espacio de memoria donde se va a ejecutar la implementación: el espacio del usuario y el espacio del kernel. En el espacio de usuario el código tiene un acceso limitado a los recursos del sistema y siempre a través de una Application Programming Interface (API) del kernel que le proporciona acceso tanto a memoria como a hardware. El código que se ejecuta en el espacio del kernel el acceso tanto a memoria como a hardware cerece de restricciones. El espacio del kernel está reservado para las funciones de más bajo nivel y confiables dado que tiene un acceso ilimitado a los recursos del sistema. Si bien una implementación en el espacio de usuario tiene un menor rendimiento que una en el espacio del kernel, se ha elegido la primera puesto que simplifica el diseño, el desarrollo, las pruebas y la portabilidad del código fuente.

El acceso a los paquetes que llegan a una interfaz del dispositivo en el espacio de usuario se hace a través de Netfilter [16], que proporciona un conjunto de puntos de enganche dentro del kernel que permiten registrar funciones *callback*. Estas funciones *callback* serán requeridas

cuando un paquete llegue al punto donde la función ha sido registrada. Hay cinco puntos de enganche: PREROUTING, por donde pasan todos los paquetes que entran en la pila de red de Linux; INPUT, por donde pasan todos los paquetes con dirección IP destino igual a cualquiera de las configuradas en el dispositivo; FORWARD, por donde pasan todos los paquetes que se reenvían a otro dispositivo; OUTPUT, por donde pasan todos los paquetes localmente generados; y finalmente, POSTROUTING, por donde pasan todos los paquetes salientes, tanto si han sido generados localmente, como si son paquetes para reenviar. La configuración de Netfilter se realiza utilizando la herramienta `iptables` o la más reciente `nftables`.

El procesado de paquetes dentro de Netfilter se realiza a través de reglas. Una regla incluye una condición de coincidencia que define a qué paquetes se le aplicará la regla y una acción que define qué se debe hacer con un paquete que cumple la condición de coincidencia. La mayor parte de las acciones se pueden reducir a dos: continuar con el procesado del paquete o bien descartarlo. Cuando a un paquete que cumple la condición se le aplica una de las dos acciones anteriores, no se siguen evaluando más reglas.

Adicionalmente a las dos acciones básicas anteriores, existe una tercera acción denominada NFQUEUE que se ajusta al diseño en el espacio de usuario de la implementación de AODVv2 propuesta en este trabajo. Así pues, cuando a un paquete se le aplica una acción NFQUEUE, se almacena en una lista enlazada que es accesible desde un código que se está ejecutando en el espacio de usuario. La comunicación entre el kernel y el espacio de usuario se realiza mediante un protocolo basado en mensajes llamado netlink [17]. Cuando un paquete se encola, el kernel notifica al software en el espacio de usuario este evento mediante un mensaje que contiene el paquete y una información adicional de asistencia. Es entonces cuando el software procesa el paquete y determina un veredicto que notifica al kernel, que actúa conforme la decisión tomada en el espacio de usuario.

La implementación de AODVv2 propuesta en este trabajo modifica tanto la Forwarding Information Base (FIB) como la Routing Information Base (RIB). Como cualquier otro protocolo de encaminamiento, AODVv2 define su propia estructura para la RIB asociada al plano de control, que utiliza para determinar la mejor ruta para un destino y que es su candidata para instalarse en la FIB. En ese caso, esa será la ruta que se utilizará para encaminar los paquetes hacia el destino.

La Fig. 3 muestra el procedimiento seguido por una aplicación Linux que envía un paquete a un nodo MANET para el que se necesita descubrir la ruta. En primer lugar, la aplicación abre un socket para enviar datos. Es entonces cuando se necesita tomar la decisión de encaminamiento. La configuración de una dirección IP en una interfaz de red instala una ruta que asume que todos los nodos de su misma red son directamente alcanzables. Sin embargo, en el caso de las redes MANET esto no es siempre cierto. De todos modos, esta ruta instalada garantiza que los paquetes destinados a la red MANET entrarán dentro del sistema Netfilter para ser procesados (i.e., la FIB tiene una ruta y el paquete IP atravesará el punto de enganche OUTPUT). En cualquier otro caso, el paquete se descartará.

Se configura una ruta en OUTPUT de tal modo que se encolan los paquetes con dirección IP destino perteneciente a la red MANET. Hay una cola distinta para cada uno de los destinos. Una cola se crea en el momento en que se detecta un destino sin cola asociada. Un hilo que se ejecuta en el espacio de usuario comprueba si existe en la RIB de AODVv2 (denominada `Local Route Set` en el RFC [5]). Si existe, se acepta el paquete y se pasa al punto de enganche POSTROUTING para finalmente ser transmitido. En caso contrario, se inicia el proceso de descubrimiento de ruta.

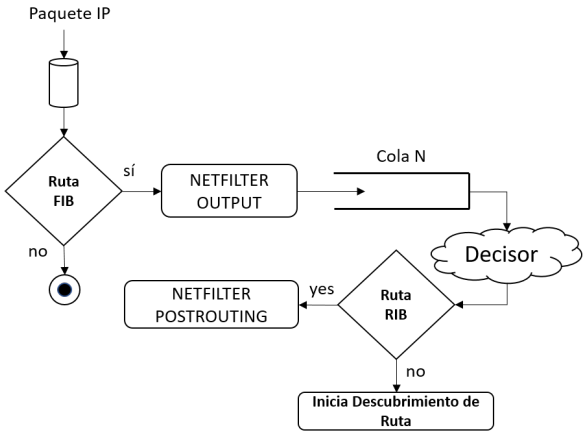


Fig. 3. Captura y procesado de un paquete IP en el espacio de usuario

Las rutas almacenadas en la RIB incluyen una única dirección destino, i.e., el tamaño del prefijo es igual al número de bits de la dirección IP destino: 32 bits en el caso de direcciones IPv4 y 128 bits en el caso de direcciones IPv6. AODVv2 define también un *Router Client Set* en cada router, i.e. un conjunto de nodos que utilizan al router como *relay*. El objetivo de este conjunto es permitir a los routers AODVv2 descubrir rutas para aplicaciones que se ejecutan en los nodos que pertenecen al *Router Client Set*, además de las que descubre para sus propias aplicaciones. En este punto, nótese que no es necesaria una ruta a la dirección destino en la FIB ya que la decisión de encaminamiento se tomará utilizando las rutas almacenadas en la RIB AODVv2. Así pues, la RIB pertenece tanto al plano de control como al de encaminamiento.

En el resto del artículo, el término nodo hace referencia a cualquier participante en la red AODVv2 y el término router hace referencia a los nodos que participan en el encaminamiento de los paquetes de datos.

A. Estructuras de Datos

La especificación del protocolo define varias estructuras de datos que tienen que ser implementados. A continuación se entra en los detalles de estas estructuras.

Como se ha mencionado anteriormente, un router AODVv2 puede descubrir rutas tanto para sus aplicaciones locales como para aquellas que se ejecutan en los nodos que pertenecen a su *Router Client Set*. Actualmente, la implementación sólo ha sido probada para aplicaciones locales, así que el *Router Client Set* solo incluye la dirección local de router y no hay una manera de añadir más nodos al conjunto. Sin embargo, la implementación de la estructura facilita la futura implementación de esta característica del protocolo.

El *Neighbor Set* es una estructura de datos que permite a los routers AODVv2 recopilar información de otros routers AODVv2 vecinos. Los enlaces con un router AODVv2 vecino se clasifican de acuerdo a su estado: CONFIRMED, HEARD y BLACKLISTED. Cuando un router AODVv2 es consciente de la presencia de un vecino, inicializa su estado a HEARD y comprueba si el enlace permite una comunicación bidireccional. Si lo permite, el router actualiza el estado a CONFIRMED. En caso contrario, el estado del router cambia a BLACKLISTED. Esto tiene su importancia ya que solo los vecinos con estado CONFIRMED pueden ser seleccionados como siguiente salto para un determinado destino.

Todas las rutas descubiertas por el protocolo se almacenan en la estructura de datos *Local Route Set*. Todas las rutas tienen una propiedad de estado con cuatro valores posibles: UNCONFIRMED, IDLE, ACTIVE e INVALID. Sólo se pueden utilizar para encaminar rutas activas. Una ruta UNCONFIRMED no se considera activa porque el enlace aún no ha sido confirmado como bidireccional. Una ruta INVALID es aquella que ha expirado o que ha detectado la caída de un enlace y, por tanto, no es una ruta válida. Las rutas con estado IDLE o ACTIVE son activas y se pueden utilizar para encaminar paquetes. Una ruta IDLE es aquella que no se ha utilizado durante un determinado tiempo configurable. Si se utiliza una ruta IDLE para encaminar un paquete, inmediantamente su estado pasa a ACTIVE.

Existen tres estructuras de datos más para completar la implementación del protocolo AODVv2. The *Multicast Message Set* se utiliza para evitar el procesado y reenvío innecesario de mensajes de control. Un mensaje RREQ se almacena en el *Multicast Message Set* y tiene asociado un temporizador de expiración (RREQ_WAIT_TIME), cuyo valor por defecto es 2. Sólo se generará un nuevo RREQ si este temporizador ha expirado.

La estructura *Route Error Set* almacena datos relacionados con direcciones inalcanzables. Finalmente, la estructura *Interface Set* almacena los detalles de todas las interfaces de red configuradas para enviar y recibir mensajes AODVv2.

La implementación se ejecuta como un demonio `systemd`, cuyo rendimiento se puede ajustar a través de los valores de un fichero de configuración. Este fichero incluye los valores de varios parámetros de AODVv2 que están agrupados en tres categorías: temporizadores, constantes del protocolo y parámetros de administración y control. Los valores de estos parámetros están inicialmente establecidos en los valores definidos por defecto en [5].

B. Mensajes del protocolo

AODVv2 define cuatro tipos de mensajes: RREQ, RREP, RERR y RREP_ACK. El primero se utiliza en el proceso de descubrimiento de rutas. Un mensaje RREP lo envía un router AODVv2 como respuesta a un mensaje RREQ destinado a cualquier dirección incluida en el *Router Client Set*. Como ya se ha dicho, en la implementación actual el *Router Client Set* solo contiene la dirección del propio router. Los mensajes RREP_ACK se utilizan para comprobar la bidireccionalidad con un vecino y pasar la ruta, en caso positivo, a estado ACTIVE. Cuando un router reenvía un mensaje RREP_ACK, a su vez envía un mensaje RREP_ACK al candidato a siguiente salto. Los mensajes RERR se generan para notificar rutas que han dejado de estar disponibles para un router.

C. Operación del protocolo

La Fig. 4 muestra el diseño del diagrama de flujo de nuestra implementación. El hilo `decisor_thread` es el responsable de interceptar los paquetes que pasan por el punto de enganche OUTPUT. Para ello se crea una regla iptables que almacena en una cola NFQUEUE todos los paquetes cuya dirección destino pertenece a la red MANET. Una vez encolados, en el espacio de usuario, el hilo utiliza la librería `libnetfilter_queue` para conectarse a la cola, la cual se caracteriza por un identificador de 16 bits (el identificador por defecto es 0), y para obtener los mensajes del kernel tan pronto como un paquete sea encolado. Esto se hace a través de una función de callback definida por el hilo `decisor_thread`. Esta función se llama cada vez que el kernel encola un paquete. La función de callback comprueba si existe una ruta activa en el *Local Route Set*. Si así ocurre, el veredicto será aceptar el paquete y este será reinyectado para continuar su camino en Netfilter pasando al punto de enganche POSTROUTING. En caso contrario, la implementación utiliza un mapa múltiple (multimap) en que cada clave corresponde a una dirección destino y tiene como valor asociado una cola en la que todos los paquetes con dicha dirección destino serán encolados y esperarán a que acabe el proceso de descubrimiento de ruta. Si no se encontrase una ruta dentro de un determinado período de expiración, todos los paquetes de la cola serían descartados y la cola, borrada.

Otro hilo, llamado `msg_thread`, es el responsable del procesado de los mensajes de control de AODVv2. Tiene un socket escuchando en la dirección multicast de enlace local 224.0.0.109, puerto 269, es decir, en la dirección multicast y puerto estándar asignada a los protocolos MANET [18]. A través de este socket se reciben y procesan los mensajes de control de AODVv2 según las especificaciones del protocolo. Así pues, el hilo `msg_thread` es el responsable del procesado de los mensajes de control de AODVv2 y el hilo `decisor_thread` es el responsable de iniciar el proceso de descubrimiento de rutas en caso de que no exista una ruta al prefijo destino. Cuando el hilo `msg_thread` descubre una ruta

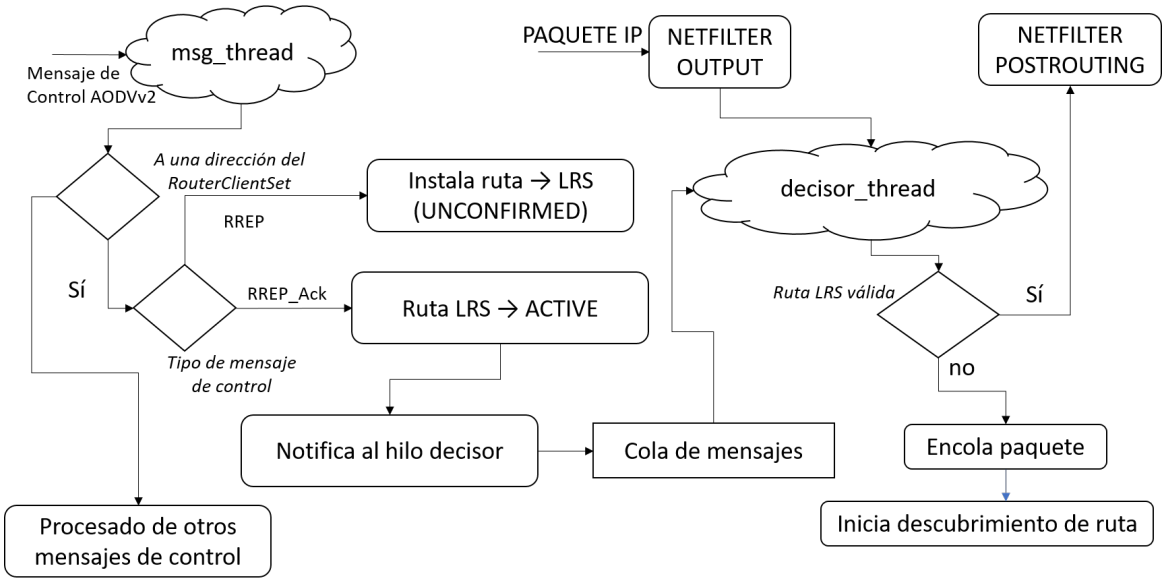


Fig. 4. Diagrama de flujo del diseño de la implementación de AODVv2. LRS se refiere a *Local Route Set*.

válida, i.e. recibe el mensaje RREP que corresponde al prefijo destino, la instala en la *Local Route Set* con estado UNCONFIRMED y envía un mensaje RREP_ACK. Al recibirse el correspondiente mensaje RREP_ACK desde el siguiente salto, la ruta pasa a tener estado VALID y el hilo `msg_thread` señaliza este evento a través de una cola de mensajes al hilo `decisor_thread` que aceptará el paquete para que continúe atravesando Netfilter.

Tanto la recepción como el procesado de los mensajes de control AODVv2 se realizan en el hilo `msg_thread`. Estas operaciones incluyen la monitorización de la disponibilidad de los siguientes saltos de una ruta, el mantenimiento del *Neighbor Set*, la transmisión de mensajes de control del protocolo en respuesta a los enviados por otros routers y el mantenimiento actualizado del *Local Route Set*.

IV. VALIDACIÓN Y PRUEBAS DE LA IMPLEMENTACIÓN

En última instancia la implementación debe ser validada y probada. En primer lugar, se ha definido una suite de pruebas completa a fin de verificar que la implementación del protocolo AODVv2 satisface el comportamiento definido en las especificaciones [5]. Finalmente, se ha ejecutado la implementación en una red sencilla con el objeto de verificar su correcto funcionamiento en dispositivos reales.

El propósito de la suite de pruebas es validar que la implementación cumple con las especificaciones del protocolo. Para ello, es preciso verificar que la generación, recepción y reenvío de cada mensaje de control AODVv2 satisface las acciones que se indican en la especificación. Para cada uno de los *casos de uso* (e.g., generación de una RREQ, reenvío de una RREQ, recepción de una RREP, etc.), se ejecuta tanto la implementación como la suite de pruebas que fuerza el *caso de uso* que queremos verificar.

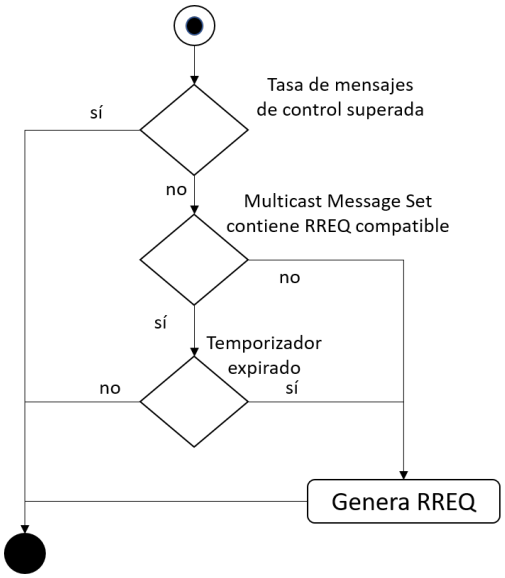


Fig. 5. Diagrama de flujo de una generación de RREQ.

Considérese, por ejemplo, el *caso de uso* de la generación de una RREQ mostrado en el diagrama de flujo de la Fig. 5. Una RREQ se genera solo si la tasa de mensajes de control no supera un cierto umbral y aplica una de las siguientes condiciones: a) el *Multicast Message Set* no contiene una RREQ compatible; o b) la contiene pero el temporizador ha expirado. Al verificar la implementación, la herramienta de pruebas prepara un estado para cada una de las posibles ramas del diagrama de flujo. En el caso que queramos comprobar si un paquete IP se envía solo si existe una ruta en la *Local Route Set* (como se explicó en la Sección III, ver Fig. 3), la herramienta instala esa ruta *antes* de intentar enviar un paquete. Con la ruta instalada, la herramienta envía un ICMP Request a través del sistema Linux y con la ayuda de una herramienta

de monitorización de tráfico de red verifica su envío. Similarmente, en el caso que queramos verificar si un paquete IP se encola cuando ya hay una RREQ compatible en curso (ver Fig. 4), la herramienta añade esta RREQ en el *Multicast Message Set* con un temporizador que aún no ha expirado y traza los elementos que hay en la cola para comprobar que el paquete está allí. Después espera hasta que el temporizador expira y vuelve a intentar enviar un ICMP Request para verificar que en esta ocasión sí se envía una RREQ.

Resumiendo, para cada *caso de uso* la herramienta genera un mensaje que satisface los requisitos y, con la ayuda de trazas y captura de los posibles mensajes generados por el router AODVv2, se verifica que la implementación es correcta.

Con la implementación comprobada localmente mediante la herramienta de verificación, se ha procedido a probar la implementación bajo condiciones más realistas. Para ello, se ha configurado una red de portátiles con arquitectura Intel x86 sobre el kernel Linux 4.4; además, para verificar su portabilidad, la implementación también ha sido probada en Raspberry Pi bajo arquitectura ARM con un kernel 4.15. En este escenario, las pruebas se han realizado sobre una sencilla red donde uno de los dispositivos, que actúa de emisor, genera tráfico ICMP destinado a otro dispositivo situado en el extremo opuesto de la red. Este escenario sirve como prueba de concepto para verificar el establecimiento de una ruta multisalto entre emisor y receptor. Todos los dispositivos tienen una interfaz de red IEEE 802.11 Wireless Network Interface Card (WNIC) que funciona en la banda de 2.4Ghz. Debido al limitado espacio físico del laboratorio donde se realizaron las pruebas, de unos 30 metros de largo, así como a la amplia cobertura de la tecnología radio utilizada, se redujo artificialmente la cobertura mediante las siguientes restricciones: primero, se configuró la tarjeta para que transmitiese a la mínima potencia posible; y, segundo, se incluyó una regla iptables en la cadena PREROUTING de cada dispositivo (ver Listado 1) para forzar la eliminación de todos los paquetes con dirección MAC origen igual a la de cada dispositivo que se desea "situar" fuera de cobertura.

Listing 1. Regla utilizada para eliminar paquetes según la dirección MAC origen

```
iptables \
  -A PREROUTING \
  -m mac --mac-source [MAC_ADDRESS] \
  -j DROP
```

La Fig. 6 muestra un escenario de pruebas con cuatro dispositivos, tanto portátiles como Raspberry Pi, etiquetados como *A*, *B*, *C* y *D*. Las flechas dobles que conectan un par de nodos representan el intercambio mutuo de datos entre dicho par de nodos. Los nodos que no están conectados por una flecha doble, no intercambian datos. En el ejemplo, el dispositivo *A* configura dos reglas iptables: una para eliminar los paquetes con dirección MAC origen igual a la dirección MAX de la interfaz del dispositivo *C* y otra

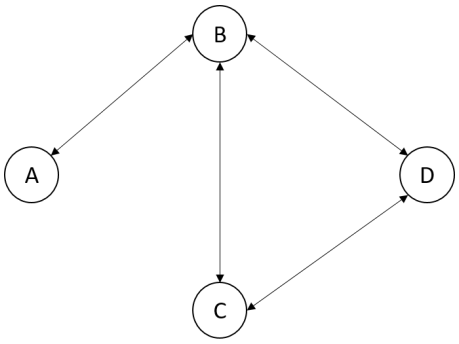


Fig. 6. RREQ Ejemplo de escenario de pruebas.

para la dirección MAX del dispositivo *D*. A su vez, los dispositivos *C* y *D* configuran sus reglas iptables con la dirección MAC del dispositivo *A* para descartar las tramas destinadas a este dispositivo.

Nótese que aunque la implementación se ha probado para las arquitecturas x86 y ARM, podría ejecutarse sin dificultades sobre otros dispositivos con sistema operativo Linux.

V. CONCLUSIONES Y TRABAJO FUTURO

El presente artículo describe una implementación del protocolo de encaminamiento AODVv2 para redes ad hoc, programado sobre el espacio de usuario del sistema operativo Linux, que aumenta la independencia del kernel y permite una fácil portabilidad sobre diferentes dispositivos. A lo largo del artículo se han discutido las decisiones de diseño de implementación y se han presentado los escenarios sobre los que se ha verificado el funcionamiento del programa. Para ello se ha creado un banco de pruebas sobre tráfico ICMP. A pesar de que existen otras implementaciones del algoritmo en la literatura sobre AODVv2, la mayoría de dichas versiones están diseñadas únicamente con fines de simulación, han quedado obsoletas o bien se circunscriben a ciertas versiones antiguas de software. La implementación propuesta en el presente artículo, que es de código abierto y mantenimiento simple, se dirige a todos los investigadores que trabajan en protocolos de encaminamiento MANET.

Actualmente, la implementación se encuentra en una etapa inicial de desarrollo, ya que precisa ser probada en escenarios más amplios y complejos para ser correctamente perfilada y, por tanto, optimizada. También contiene trazas de desarrollo con el fin de facilitar el trabajo de depuración y ajuste. Estos fragmentos de código, que se pueden desactivar en términos de funcionalidad, se eliminarán cuando se considere que el código está listo para su libre distribución.

Como se señala en la Sección III, la funcionalidad *Router Client Set* queda pendiente de implementar. Además, aunque no resulta necesario para el proceso actual de prueba y validación, está previsto ampliar la configuración de la aplicación para permitir múltiples interfaces inalámbricas en cada dispositivo. Una vez haya concluido esta labor de creación de perfiles, se llevará a cabo una prueba más ambiciosa, con varios dispositivos Raspberry

Pi distribuidos a lo largo de un mismo edificio, con el fin de evaluar el funcionamiento de la implementación del protocolo en condiciones más realistas. Esto supondrá el diseño de encaminamientos más complejos así como el empleo de patrones de tráfico más realistas que el tráfico ICMP empleado en el presente trabajo.

REFERENCIAS

[1] A. Mishra, S. Singh, A. Tripathi, "Comparison of Manet Routing Protocol," International Journal of Computer Science and Mobile Computing, vol. 8, pp. 67-74, 2019.

[2] P. Shailaja, C.V. Guru Rao, A.Nagaraju, "A Parametric Oriented Research on Routing Algorithms in Mobile Adhoc Networks," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 1, November 2019.

[3] I. Snigdh, D. Gosain, (2015). Analysis of scalability for Routing Protocols in Wireless Sensor Networks. Optik - International Journal for Light and Electron Optics. 127. 10.1016/j.ijleo.2015.11.077.

[4] Perkins, C., Belding-Royer, E., and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC 3561, DOI 10.17487/RFC3561, July 2003, https://www.rfc-editor.org/info/rfc3561.

[5] C. Perkins, S. Ratliff, J. Dowdell, L. Steenbrink, and V. Mercieca, "Ad Hoc On-demand Distance Vector Version 2 (AODVv2) Routing," Work in Progress, draft-perkins-manet-aodvv2, May 2016.

[6] T. Clausen, C. Dearlove, J. Dean, and C. Adjih, "Generalized Mobile Ad Hoc Network (MANET) Packet/Message Format", RFC 5444, DOI 10.17487/RFC5444, February 2009.

[7] T. Kumar-Saini and S.C. Sharma, "Recent advancements, review analysis, and extensions of the AODV with the illustration of the applied concept," Ad Hoc Networks, 103, 2020, https://doi.org/10.1016/j.adhoc.2020.102148.

[8] P. Rajankumar, P. Nimisha and P. Kamboj, "A comparative study and simulation of AODV MANET routing protocol in NS2 & NS3," 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2014, pp. 889-894, doi: 10.1109/IndiaCom.2014.6828091.

[9] C. Sommer, I. Wagner, F. Dressler, "A simulation model of DYMO for ad hoc routing in OMNeT++," Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops, SimuTools 2008, Marseille, France, March 3-7, 2008.

[10] AODV-UU source from GitHub, April 13, 2011. https://github.com/erimatnor/aodv-uu (last accessed on March 2021)

[11] W. Backes and J. Cordasco, "MoteAODV – An AODV Implementation for TinyOS 2.0.," Information Security Theory and Practices. Security and Privacy of Pervasive Systems and Smart Devices. WISTP 2010. Lecture Notes in Computer Science, 6033. Springer, doi: 10.1007/978-3-642-12368-9_11.

[12] E. Gaona-García, S. Palechor-Mopán, L. Murcia-Sierra, P. Gaona-García, "Implementation of the AODV Routing Protocol for Message Notification in a Wireless Sensor Microgrid," 5th Workshop on Engineering Applications, WEA 2018, Medellín, Colombia, October 17-19, 2018, Proceedings, Part II. 10.1007/978-3-030-00353-1_32.

[13] J. S. Awati, S. A. Patil and M. R. Patil, "Implementation of AODV Routing Protocol of Wireless Sensor Network in Agriculture," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 2018, pp. 1-9, doi: 10.1109/ICOEI.2018.8553730.

[14] S. Jung, B. Kim, K. Kim, B. Roh and J. Ham, "Implementation of AODV-UU on Linux 4.15 Kernel," 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW), Monterey, CA, USA, 2019, pp. 160-161, doi: 10.1109/MASSW.2019.00039.

[15] P. García López, R. García Tinedo, and J. M. Banús Alsina, "Moving routing protocols to the user space in MANET middleware," 2010 Journal of Network and Computer Applications, pp. 588-602.

[16] http://netfilter.org. Visited the 21th December 2020.

[17] Salim, J., Khosravi, H., Kleen, A., and A. Kuznetsov, "Linux Netlink as an IP Services Protocol", RFC 3549, DOI 10.17487/RFC3549, July 2003, ¡https://www.rfc-editor.org/info/rfc3549¿.

[18] Chakeres, I., "IANA Allocations for Mobile Ad Hoc Network (MANET) Protocols", RFC 5498, DOI 10.17487/RFC5498, March 2009.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Análisis teórico para la mejora del rendimiento en redes LoRa

Jose M. Jimenez, Laura García, Sandra Sendra, Jaime Lloret.
Instituto de Investigación para la Gestión Integrada de Zonas Costeras,
Universitat Politècnica de València
46730 Grau de Gandia, Valencia, Spain.
jojiher@dcom.upv.es, laugarg2@teleco.upv.es, sansenco@upv.es, jlloret@dcom.upv.es

Las tecnologías Low Power Wide Area Network (LPWAN) no celulares están siendo cada día más estudiadas y aplicadas en múltiples entornos. LoRa (Long Range) es una de las tecnologías de comunicación LPWAN que se está implementando con mayor frecuencia, debido principalmente a sus características, entre las que podemos destacar su uso de bandas ISM, el bajo consumo de energía y largo alcance. Por contra, su uso presenta una serie de desventajas que deben ser solucionadas; de entre ellas podemos destacar una capacidad de cobertura global limitada (frecuencias de trabajo diferentes por continentes), una tasa de transmisión lenta, una capacidad de carga útil baja, problemas seguridad, etc. Nuestro trabajo presenta un estudio de LoRa, con el fin de determinar si es posible enviar otro tipo de información como imágenes y vídeo, calculando el porcentaje de carga útil máxima. Para ello, partiendo de diferentes imágenes estándares, calcularemos de forma teórica la distancia, número de paquetes, Time on Air (ToA) y energía, empleados para la transmisión, y definidos por la propia tecnología, que nos sirven de base para un realizar en el futuro una adecuada selección de dispositivos y tipos de archivos para la transmisión de imágenes. Con ello ampliaríamos las posibilidades de uso de esta tecnología en otros ámbitos.

***Palabras Clave*- LoRa, Internet de las Cosas (IoT), Redes de sensores.**

I. INTRODUCCIÓN

El Internet de las Cosas (IoT) forma en actualidad parte de nuestra vida corriente. Cada vez estamos más rodeados de dispositivos y máquinas, que están conectadas mediante diferentes tipos de redes de comunicación. Las tres características principales que caracterizan a las redes inalámbricas son su tasa de transmisión o data rate, su alcance y su consumo energético. En los últimos años se ha producido un gran despliegue de redes basadas en tecnologías Low Power Wide Area Network (LPWAN). Generalmente, estas tecnologías, permiten la transmisión de datos a grandes distancias con un reducido consumo energético. Por estas características, se hace muy atractivo su aplicación en los dispositivos que se emplean en IoT. Las tecnologías más conocidas, que nos permiten un alcance medio–largo, son LoRa, SigFox, LTE-M (Long Term Evolution-M) o NB-IOT (Narrow Band Internet of Things). Las dos últimas tecnologías LTE-M y NB-IOT son utilizadas en redes celulares, por operadores de red móvil. La tecnología SigFox está controlada por la propia empresa SigFox y es necesario tener un plan de subscripción para cada dispositivo que se quiera conectar. Por lo tanto, desde el punto de vista económico, la conexión a través de la tecnología LoRa, es la más rentable para el usuario, pues solamente necesita adquirir los dispositivos y conectarlos. Aunque también hay operadores, como Orange, que ofrece redes LoRa en algunas áreas.

Lora opera en el espectro de frecuencias industrial, científico y médico (ISM). Por lo cual, podemos crear redes e interconectar los equipos, respetando las bandas permitidas sin necesidad de pedir licencias de instalación.

En muchos de los ámbitos donde se aplica IoT, la cantidad de datos que es necesario transmitir es pequeña, y a la vez, aunque se pueden hacer observaciones en tiempo real, no es necesario un envío continuo de los datos observados. Sería el caso de aplicaciones en el entorno de Smart Agriculture, o en entornos donde se observa un estado y solo nos interesa conocer su cambio de estado, entre otros muchos casos.

Existen otros ámbitos donde la cantidad de datos a transmitir es mucho mayor, por ejemplo, cuando se intenta transmitir imágenes, en entornos como puede ser el de la vídeo vigilancia, en el que se suelen enviar imágenes fijas o en movimiento.

En nuestro trabajo estudiaremos la capacidad de transmisión de imágenes o vídeos, empleando la tecnología LoRa. Partiendo de diferentes imágenes estándares, calcularemos de forma teórica la distancia, número de paquetes, Time on Air (ToA) y energía, empleados para la transmisión, según los define el fabricante. De esta forma, obtendremos unos datos iniciales que emplearemos para continuar nuestras investigaciones, y que permitan seleccionar los formatos de compresión de imágenes y los nodos más adecuados,

para ser empleados en entornos reales. cuando sea necesaria la transmisión de imágenes.

El resto del documento se estructura de la siguiente forma. En la sección 2 hacemos una descripción de la tecnología LoRa. En la sección 3 presentamos el estudio para la realización de la transmisión, con las capacidades de carga óptima. En la sección 4 presentaremos la discusión sobre los resultados observados. En la sección 5 presentaremos las conclusiones y trabajos futuros.

II. DESCRIPCIÓN LORA

En este apartado se presenta una visión general de la tecnología LoRa y sus principales características.

Long Range (LoRa) es una tecnología inalámbrica en la que un transmisor de baja potencia envía pequeños paquetes de datos (entre 0,3 kbps y 5,5 kbps) a un receptor, generalmente a larga distancia. Estas características hacen que LoRa se encuentre en la categoría de redes LPWAN [1]. Aunque el rango de cobertura de LoRa es amplio, depende en gran medida del medio ambiente y de los materiales de construcción que lo componen. El rango de los alrededores rurales es de aproximadamente 20 km. Sin embargo, en entornos urbanos, se reduce a 5 km. No obstante, algunas pruebas confirman que los datos se recibieron desde una distancia de cientos de kilómetros con una visión directa adecuada y sin obstáculos en el área de Fresnel entre los dispositivos [2] [3].

La tecnología LoRa Wireless fue desarrollada por la empresa francesa Cycleo. En 2012 la empresa estadounidense Semtech adquirió Cycleo, que ahora posee la patente de la parte de radio y la modulación LoRa y cuyos códigos están cerrados. Semtech proporciona licencias de propiedad intelectual a otras empresas, especialmente a fabricantes de componentes como HopeRF, Microchip, etc.

Por otro lado, los protocolos LoRaWAN son abiertos y definidos por LoRa Alliance, una organización sin fines de lucro fundada en 2015 con más de 500 empresas colaboradoras (IBM, Microchip, Orange, Cisco, etc.) creadas con el compromiso de permitir e incentivar implementaciones a gran escala de dispositivos LPWAN IoT que implementan sus estándares.

A. Resumen de las ventajas de LoRa

Las características más importantes de LoRa con respecto a su implementación son [4]:

- Transmisión ortogonal de paquetes, utilizando diferentes SF (de 6 a 12) simultáneamente, sin colisiones de datos ni relegación en la funcionalidad, disminuyendo el tiempo en el aire (TOA).
- Transmisión de largo alcance, donde la cobertura al aire libre puede extenderse a decenas de kilómetros, mientras que la cobertura en interior se acerca al kilómetro y la cobertura en el uso con múltiples saltos se acerca a varios kilómetros.
- Bajo precio del desarrollo de la red, porque los nodos pueden conectarse directamente al sumidero sin rutas mediadoras.
- Utiliza una variedad de tasas de datos y es multicanal, para recibir datos de una gran cantidad de nodos de red.
- Bajo consumo de energía para la conservación de la batería.
- Alta capacidad de penetración de señales de radio.
- Funcionamiento en las bandas sin licencia.

B. Modulación LoRa

La modulación es la forma en que se codifica la información (digital o analógica) en la señal portadora que será la encargada de transmitirla. La modulación LoRa se basa en CSS (Chirp Spread Spectrum), que es una técnica de espectro ensanchado que utiliza pulsos de chirp de modulación de frecuencia lineal, con gran ancho de banda, para codificar la información. Un pulso de chirp no es más que una señal sinusoidal en la que la frecuencia aumenta (up-chirp) o disminuye (down-chirp) con el tiempo. Un chirp determina un símbolo.

Los chirps cambian de frecuencia durante un cierto período de tiempo, llamado tiempo de símbolo (Ts). Estos "saltos" de frecuencia determinan cómo se codifica la información. El número de bits que se pueden codificar en cada símbolo viene dado por el factor de expansión (SF), por lo que un símbolo puede tener valores 2SF para saltar. Estos valores se denominan chirps (por ejemplo, si SF es 7, el número de bits que el símbolo puede codificar también es 7 y tiene 27 chirps). LoRa admite un rango de valores enteros de SF entre 7 y 12 [5].

La velocidad de símbolo Rs se define, como se indica en [6], como:

$$R_s(symbol/s) = \frac{BW}{2^{SF}} \quad (1)$$

Dado que la tasa de chirp es constante para un ancho de banda dado (Rc = BW), (1) se puede reescribir como (2):

$$R_s = \frac{R_c}{2^{SF}} \quad (2)$$

La tasa de bits (Rb) se define como:

$$R_b(bps) = \frac{SF\ BW}{2^{SF}}\ \frac{4}{4+CR} \quad (3)$$

donde CR (Tasa de codificación) es la proporción de bits no redundantes para la corrección de errores hacia adelante (FEC). La modulación LoRa permite diferentes valores para esto:

$$CR = \frac{4}{4+n} \quad (4)$$

donde n = 1, 2, 3 y 4, por lo que los valores permitidos son 4/5, 4/6, 4/7 y 4/8.

La duración (en segundos) se puede calcular como

$$\mathrm{Tc\ (s)} = \frac{1}{BW} \quad (5)$$

ya que, como dijimos anteriormente, Rc = BW. La duración de los símbolos se define entonces por:

$$T_s(s) = \frac{2^{SF}}{BW} \quad (6)$$

Como muestran las ecuaciones, cuanto mayor es el valor SF, menor es la tasa de bits y mayor es el tiempo del símbolo. Por lo tanto, aumentará el tiempo de transmisión del mensaje o Time on Air (ToA) y el consumo de energía en el transmisor del dispositivo. Por otro lado, el rango de cobertura será mayor a un valor SF más alto debido a la mayor robustez resultante frente al ruido [7].

El Time on Air (ToA) es la cantidad de tiempo antes de que un receptor reciba una señal de un remitente.

Con respecto al uso de diferentes canales, las señales recibidas con diferentes SF generalmente se consideran puramente ortogonales. Sin embargo, no es cierto en determinadas condiciones de nivel de potencia. Los parámetros del nodo final (SF y potencia de transmisión) se pueden ajustar en función de la distancia desde la puerta de enlace y ofrece la posibilidad de ejecutar redes con múltiples puertas de enlace [8] [9] [10].

Además, otra característica de la modulación LoRa es su inmunidad frente al efecto Doppler. El desplazamiento causado por el efecto Doppler provoca un pequeño cambio de frecuencia a la señal modulada que apenas afecta a la señal de banda base en el dominio del tiempo.

C. Regulación

LoRa opera en la banda sin una licencia ISM que esté disponible en todo el mundo. En Europa, las bandas de frecuencia con licencia para este uso son las bandas EU433 (433,05–434,79 MHz) y EU863-870 (863–870 MHz). En los EE. UU., Es la banda US902-928 (902–928 MHz). Esta banda tiene la principal ventaja de ser de uso gratuito, sin licencia, pero, debido a su gran uso, hay mucha interferencia y la tasa de transmisión es baja. Estas bandas están reguladas por diferentes organizaciones. En Europa, está regulado por el Instituto Europeo de Telecomunicaciones (ETSI). En EE. UU., Por la Comisión Federal de Comunicaciones (FCC). La mayoría de los países copian las reglas estandarizadas por estas agencias a excepción de algunos países como Japón y Corea del Sur que tienen sus propias agencias.

D. Limitaciones de la transmisión de imágenes a través de LoRa

La tecnología LoRa restringe el ciclo de trabajo al 1% (es decir, 36 s / h) y aplica al tiempo total de transmisión, lo que significa que solo se pueden transmitir datos durante 36 s cada hora de tiempo de transmisión [11]. Por ejemplo, si hay un ToA de 500 ms, se puede enviar un mensaje de nuevo en 99 × 500 = 49,500 ms, 49,5 s.

Esta limitación hace que la transferencia de datos desde dispositivos como sensores de imagen sea muy compleja, ya que se requiere una gran velocidad de bits para la comunicación de datos de imagen. Por lo tanto, aunque LoRa tiene la capacidad de combinar las características de las redes LPWA y las redes WSN, la transferencia de datos multimedia a través de LoRa es un desafío debido a la limitación de transmisión de datos.

III. Estudio teórico

En esta sección se muestran los resultados del estudio teórico de la capacidad de LoRa para enviar imágenes. Para ello, se han evaluado la distancia máxima, el número de paquetes, el ToA y el consumo energético.

Los formatos de imagen estudiados y sus características se detallan en la Tabla 1. Los formatos elegidos son usados para la transmisión de imágenes y vídeo, desde los más pequeños (CGA 320X200) hasta los mayores (HD 1080 o QXGA), con diferentes relaciones de aspecto (8:5, 4:3, 16:9). Consideramos que serán los formatos más empleados por los dispositivos de captura de imagen, y empleados posteriormente en la transmisión. Se ha realizado un estudio teórico para cada uno de los formatos y sus dos profundidades de color empleando las fórmulas detalladas en [12].

Tabla I. Formatos de imagen

Estándar	Relación Aspecto	Alto (Px)	Ancho (pixel)	Profundidad de Color	Tamaño (bytes)
CGA	8:5	320	200	1	8000
	8:5	320	200	24	192000
WXGA	8:5	1280	800	1	128000
	8:5	1280	800	24	3072000
WUXGA	8:5	1920	1200	1	288000
	8:5	1920	1200	24	6912000
QVGA	4:3	320	240	1	9600
	4:3	320	240	24	230400
VGA	4:3	640	480	1	38400
	4:3	640	480	24	921600
SVGA	4:3	800	600	1	60000
	4:3	800	600	24	1440000
XGA	4:3	1024	768	1	98304
	4:3	1024	768	24	2359296
QXGA	4:3	2048	1536	1	393216
	4:3	2048	1536	24	9437184
FWVGA	16:9	854	480	1	51240
	16:9	854	480	24	1229760
HD720	16:9	1280	720	1	115200
	16:9	1280	720	24	2764800
HD1080	16:9	1920	1080	1	259200
	16:9	1920	1080	24	6220800

A. Distancia

La distancia máxima teórica a la que pueden transmitir los dispositivos LoRa depende de la configuración del SF, el ancho de banda y de la frecuencia a la que se esté transmitiendo. Para la frecuencia 868MHz, la cual se corresponde con la normativa europea, los resultados de la distancia máxima se muestran en la Figura 1. Como se puede apreciar, a mayor SF, mayor distancia. Asimismo, a mayor ancho de banda, menor distancia máxima. Se han realizado también los cálculos para las frecuencias de 250 y 500 KHz, tal como se precia en la Figura 1.

B. Número de paquetes enviados

Los diferentes estándares de imagen dan lugar a ficheros de diversos tamaños. Esto influye en el número de paquetes que se han de enviar para transmitir la imagen en su totalidad. En la Tabla 2 se muestran los resultados. Como se puede apreciar en la Figura 2, el formato QXGA es el que mayor cantidad de paquetes necesita enviar, con 38997 paquetes para el formato con profundidad de color de 24 bits. Considerando la gran diferencia entre enviar imágenes en blanco y negro y enviar imágenes a color, se debería considerar la realización de un procesado previo de las imágenes antes de enviarlas inalámbricamente.

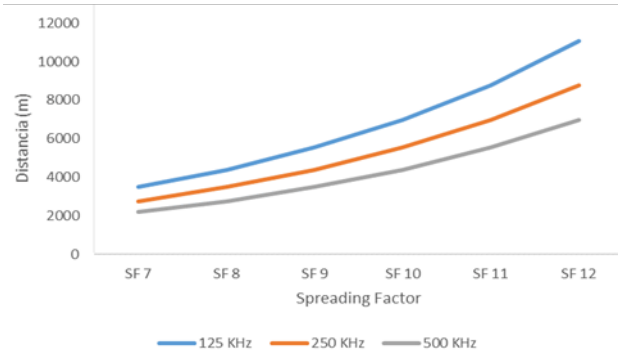


Fig. 1. Distancia máxima teórica.

Tabla II. NÚMERO DE PAQUETES ENVIADOS

ESTÁNDAR	PROFUNDIDAD DE COLOR (BITS)	NÚMERO DE PAQUETES
CGA	*1*	*34*
	24	*794*
WXGA	*1*	*529*
	24	*12695*
WUXGA	*1*	*1191*
	24	*28562*
QVGA	*1*	*40*
	24	*953*
VGA	*1*	*159*
	24	*3809*
SVGA	*1*	*248*
	24	*5951*
XGA	*1*	*4062*
	24	*9750*
QXGA	*1*	*1625*
	24	*38997*
FWVGA	*1*	*212*
	24	*5082*
HD 720	*1*	*477*
	24	*11425*
HD 1080	*1*	*1072*
	24	*25706*

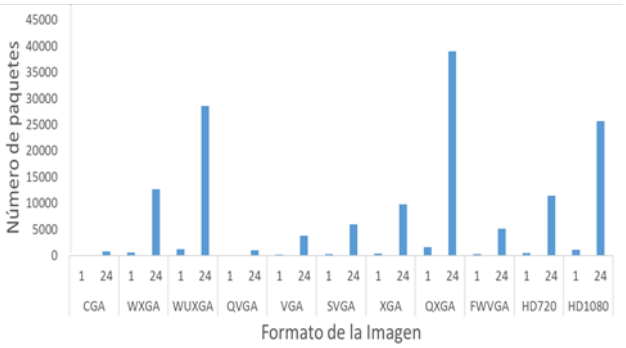


Fig. 2. Número de paquetes a enviar para cada uno de los formatos de imagen.

C. *Time on Air*

El tamaño de la imagen también afecta al ToA. Tal como puede verse en las Figuras 3 and 4, el ToA es mayor para el ancho de banda de 125 KHz. Sin embargo, este ancho de banda es el que se debe usar en gran parte de las áreas debido a restricciones locales. Asimismo, a mayor SF, mayor ToA. Es por ello por lo que se debe considerar las necesidades de la aplicación para determinar si se prioriza la distancia o el ToA a la hora de seleccionar los parámetros de LoRa. El formato de imagen QXGA es el que presenta mayor ToA. Con un tiempo de transmisión de 10 minutos aproximadamente en el caso de la imagen de blanco y negro y 4 horas aproximadamente para la imagen a color. Por otra parte, el formato que menos tiempo requiere para envía la imagen es CGA con 12.14 segundos. En las Tablas 3, 4 y 5 se muestran los valores numéricos para cada imagen.

D. *Energía*

Finalmente, en las Figuras 5 y 6 se muestra el consumo energético para las imágenes en blanco y negro y las imágenes en color respectivamente. Como puede apreciarse, el formato de imagen con mayor consumo energético es el QXGA con 1052433 mJ para las imágenes en blanco y negro y 25257726 mJ para las imágenes a color. Los resultados de consumo energético para cada uno de los formatos de imagen están disponibles en las Tablas 3, 4 y 5.

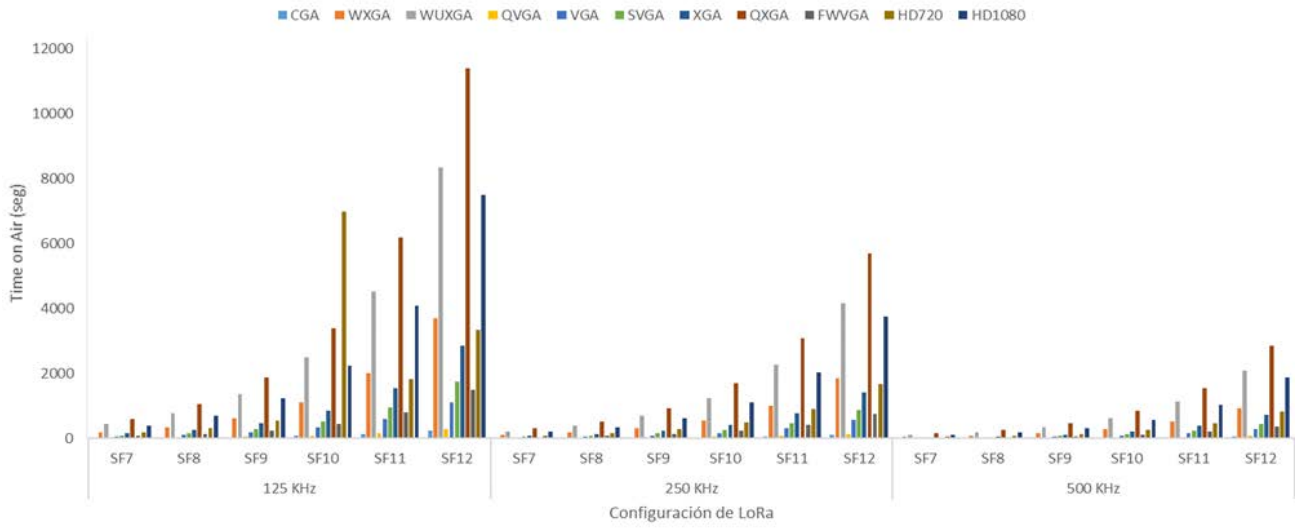


Fig. 3. ToA para las imágenes en blanco y negro.

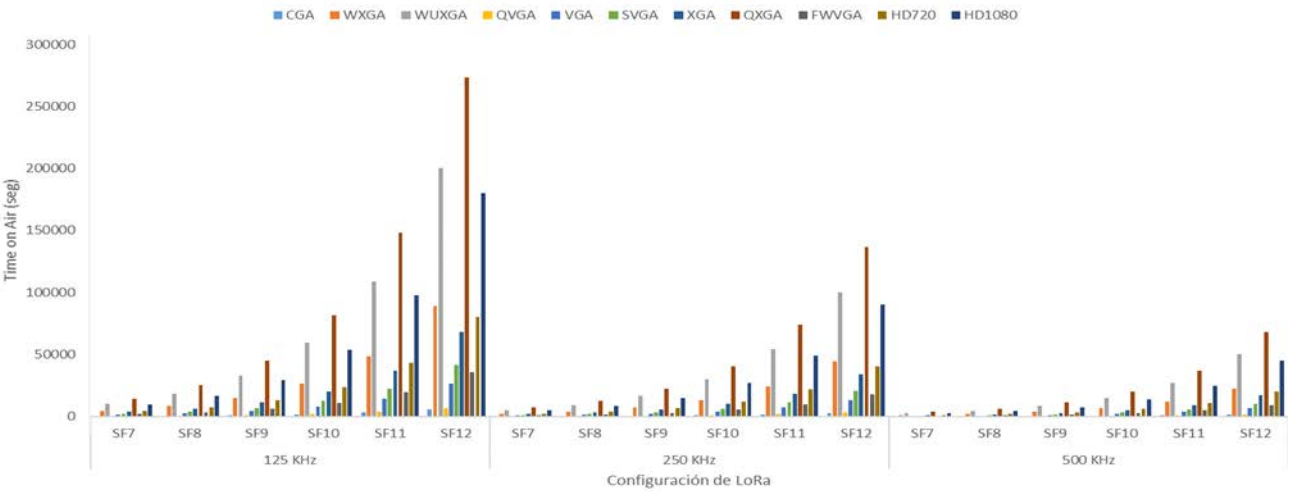


Fig. 4. ToA para las figuras con una profundidad de color de 24 bits.

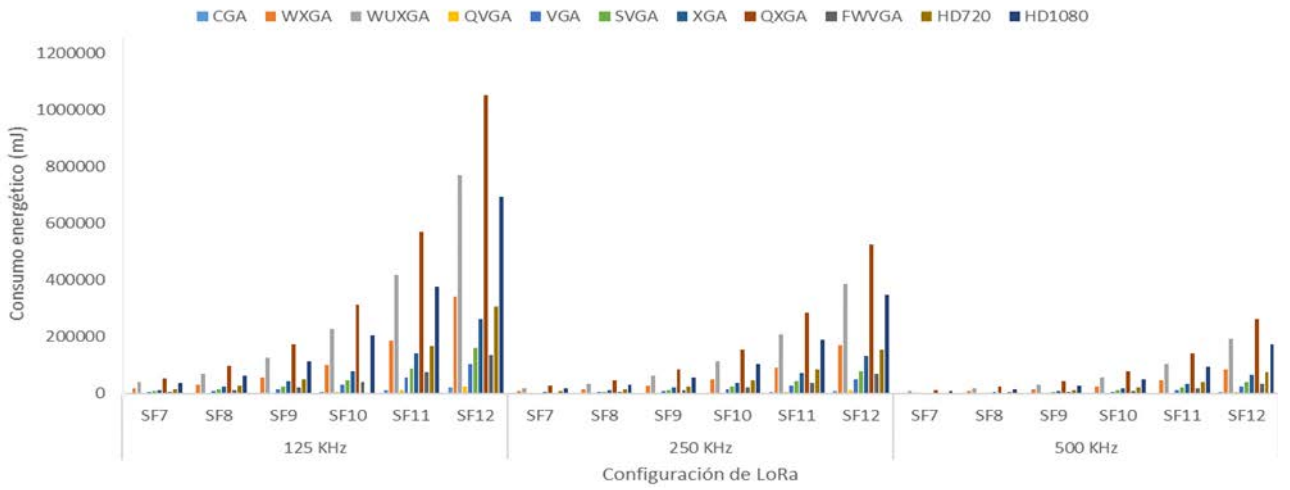


Fig. 5. Consumo energético para las imágenes en blanco y negro.

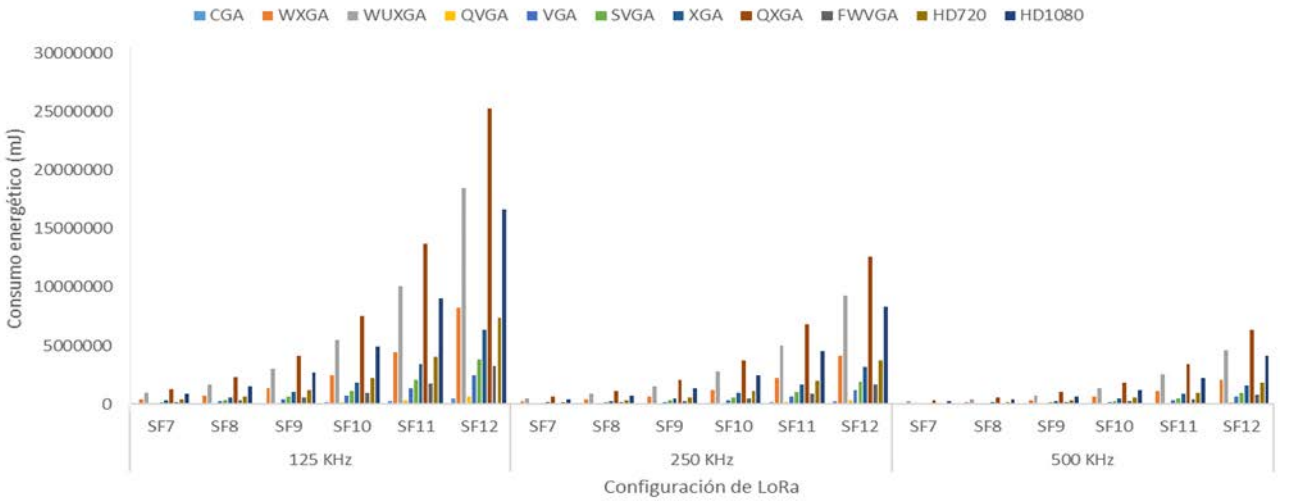


Fig. 6. Consumo energético para las figuras con una profundidad de color de 24 bits.

TABLA III. VALORES DE TOA (SEG) Y CONSUMO DE ENERGÍA (MJ) PARA CADA FORMATO DE IMAGEN Y EL ANCHO DE BANDA DE 125 KHZ

		ToA						Consumo energético en mJ					
Estandar	Profundidad de Color	SF7	SF8	SF9	SF10	SF11	SF12	SF7	SF8	SF9	SF10	SF11	SF12
CGA	1	12.14	21.36	38.15	69.05	126.15	232.47	1122	1974	3525	6381	11656	21480
	24	290.84	511.47	913.69	1652.65	3019.32	5561.86	26874	47260	84425	152705	278985	513916
WXGA	1	193.89	340.98	609.13	1101.77	2012.83	3707.97	17916	31506	56284	101803	185985	342616
	24	4653.16	8182.84	14618.04	26439.9	48304.16	88982.22	429952	756094	1350707	2443047	4463304	8221957
WUXGA	1	436.26	767.19	1370.53	2478.92	4528.83	8342.66	40310	70888	126637	229052	418464	770862
	24	10469.56	18411.31	32890.46	59489.46	108683.8	200208.96	967388	1701205	3039079	5496827	10042383	18499308
QVGA	1	14.56	25.61	45.74	82.76	151.19	278.46	1345	2366	4227	7647	13970	25730
	24	349.01	613.76	1096.45	1983.18	3623.11	6674.26	32249	56711	101312	183246	334776	616702
VGA	1	58.18	102.32	182.79	330.64	604.01	1112.66	5376	9454	16890	30551	55810	102810
	24	1395.96	2454.88	4385.48	7932.09	14491.46	26695.05	128987	226831	405218	732926	1339010	2466623
SVGA	1	90.9	159.85	285.56	516.51	943.65	1738.21	8399	14770	26385	47726	87194	160611
	24	2181.18	3835.73	6852.25	12393.78	22642.73	41710.64	201541	354421	633148	1145185	2092189	3854063
XGA	1	148.92	261.9	467.86	846.22	1546.06	2847.99	13761	24199	43230	78191	142856	263154
	24	3573.63	6284.43	11226.69	20305.88	37097.64	68338.5	330204	580682	1037346	1876264	3427822	6314477
QXGA	1	595.61	1047.43	1871.15	3384.4	6183.06	11389.97	55035	96782	172894	312718	571315	1052433
	24	14294.44	25137.57	44906.44	81222.91	148389.58	273352.01	1320806	2322712	4149355	7504997	13711197	25257726
FWVGA	1	77.63	136.52	243.88	441.14	805.93	1484.55	7173	12614	22534	40761	74468	137173
	24	1862.73	3275.71	5851.82	10584.29	19336.85	35620.86	172116	302676	540708	977988	1786725	3291368
HD720	1	174.52	306.91	548.28	6977.95	1811.76	3337.46	16125	28358	50661	477	167407	308381
	24	4187.84	7364.55	13156.23	23795.87	43473.62	80083.82	386956	680484	1215635	2198738	4016963	7399745
HD1080	1	392.63	690.47	1233.49	2231.05	4076.01	7508.46	36279	63800	113974	206149	376623	693782
	24	9422.61	16570.19	29601.43	53540.55	97815.45	180188.17	870649	1531085	2735172	4947147	9038148	16649387

TABLA IV. VALORES DE TOA (SEG) Y CONSUMO DE ENERGÍA (MJ) PARA CADA FORMATO DE IMAGEN Y EL ANCHO DE BANDA DE 250 KHZ

		ToA						Consumo energético en mJ					
Estandar	Profundidad de Color	SF7	SF8	SF9	SF10	SF11	SF12	SF7	SF8	SF9	SF10	SF11	SF12
CGA	1	6.07	10.68	19.08	34.53	63.07	116.24	561	987	1763	3190	5828	10740
	24	145.42	255.73	456.85	826.32	1509.66	2780.93	13437	23630	42213	76352	139492	256958
WXGA	1	96.95	170.49	304.56	550.88	1006.41	1853.98	8958	15753	28142	50902	92993	171308
	24	2326.58	4091.42	7309.02	13219.95	24152.08	44491.11	214976	378047	675353	1221523	2231652	4110978
WUXGA	1	218.13	383.59	685.27	1239.46	2264.41	4171.33	20155	35444	63319	114526	209232	385431
	24	5234.78	9205.66	16445.23	29744.73	54341.9	100104.48	483694	850603	1519539	2748413	5021191	9249654
QVGA	1	7.28	12.8	22.87	41.38	75.6	139.23	673	1183	2113	3824	6985	12865
	24	174.51	306.88	548.22	991.59	1811.56	3337.13	16124	28356	50656	91623	167388	308351
VGA	1	29.09	51.16	91.39	165.32	302	556.33	2688	4727	8445	15275	27905	51405
	24	697.98	1227.44	2192.74	3966.05	7245.73	13347.52	64493	113416	202609	366463	669505	1233311
SVGA	1	45.45	79.92	142.78	258.26	471.83	869.11	4199	7385	13193	23863	43597	80305
	24	1090.59	1917.86	3426.13	6196.89	11321.37	20855.32	100770	177211	316574	572593	1046094	1927032
XGA	1	74.46	130.95	233.93	423.11	773.03	1423.99	6880	12100	21615	39096	71428	131577
	24	1786.82	3142.22	5613.34	10152.94	18548.82	34169.25	165102	290341	518673	938132	1713911	3157239
QXGA	1	297.81	523.71	935.58	1692.2	3091.53	5694.98	27517	48391	86447	156359	285658	526217
	24	7147.22	12568.79	22453.22	40611.46	74194.79	136676	660403	1161356	2074678	3752499	6855599	12628863
FWVGA	1	38.81	68.26	121.94	220.57	402.96	742.28	3586	6307	11267	20381	37234	68586
	24	931.36	1637.85	2925.91	5292.14	9668.42	17810.43	86058	151338	270354	488994	893362	1645684
HD720	1	87.26	153.45	274.14	495.85	905.88	1668.73	8063	14179	25330	45816	83703	154191
	24	2093.92	3682.27	6578.11	11897.93	21736.81	40041.91	193478	340242	607818	1099369	2008481	3699873
HD1080	1	196.32	345.24	616.74	1115.53	2038.01	3754.23	18140	31900	56987	103075	188312	346891
	24	4711.3	8285.09	14800.71	26770.27	48907.73	90094.08	435325	765543	1367586	2473573	4519074	8324693

TABLA V. VALORES DE TOA (SEG) Y CONSUMO DE ENERGÍA (MJ) PARA CADA FORMATO DE IMAGEN Y EL ANCHO DE BANDA DE 500 KHZ

		ToA						Consumo energético en mJ					
Estandar	Profundidad de Color	SF7	SF8	SF9	SF10	SF11	SF12	SF7	SF8	SF9	SF10	SF11	SF12
CGA	1	3.04	5.34	9.54	17.26	31.54	58.12	281	493	881	1595	2914	5370
	24	72.71	127.87	228.42	413.16	754.83	1390.47	6718	11815	21106	38176	69746	128479
WXGA	1	48.47	85.24	152.28	275.44	503.21	926.99	4479	7877	14071	25451	46496	85654
	24	1163.29	2045.71	3654.51	6609.97	12076.04	22245.55	107488	189024	337677	610762	1115826	2055489
WUXGA	1	109.06	191.8	342.63	619.73	1132.21	2085.66	10078	17722	31659	57263	104616	192715
	24	2617.39	4602.83	8222.62	14872.37	27170.95	50052.24	241847	425301	759770	1374207	2510596	4624827
QVGA	1	3.64	6.4	11.44	20.69	37.8	69.62	336	592	1057	1912	3493	6432
	24	87.25	153.44	274.11	495.79	905.78	1668.56	8062	14178	25328	45811	83694	154175
VGA	1	14.54	25.58	45.7	82.66	151	278.17	1344	2364	4222	7638	13953	25702
	24	348.99	613.72	1096.37	1983.02	3622.86	6673.76	32247	56708	101305	183231	334753	616656
SVGA	1	22.72	39.96	71.39	129.13	235.91	434.55	2100	3692	6596	11931	21798	40153
	24	545.29	958.93	1713.06	3098.45	5660.68	10427.66	50385	88605	158287	286296	523047	963516
XGA	1	37.23	65.47	116.97	211.56	386.51	712	3440	6050	10808	19548	35714	65789
	24	893.41	1571.11	2806.67	5076.47	9274.41	17084.62	82551	145170	259336	469066	856956	1578619
QXGA	1	148.9	261.86	467.79	846.1	1545.77	2847.49	13759	24196	43224	78180	142829	263108
	24	3573.61	6284.39	11226.61	20305.73	37097.4	68338	330202	580678	1037339	1876249	3427799	6314431
FWVGA	1	19.41	34.13	60.97	110.28	201.48	371.14	1793	3154	5634	10190	18617	34293
	24	465.68	818.93	1462.96	2646.07	4834.21	8905.22	43029	75669	135177	244497	446681	822842
HD720	1	43.63	76.73	137.07	247.92	452.94	834.37	4031	7090	12665	22908	41852	77095
	24	1046.96	1841.14	3289.06	5948.97	10868.41	20020.96	96739	170121	303909	549685	1004241	1849936
HD1080	1	98.16	172.62	308.37	557.76	1019	1877.11	9070	15950	28494	51537	94156	173445
	24	2355.65	4142.55	7400.36	13385.14	24453.86	45047.04	217662	382771	683793	1236787	2259537	4162347

IV. DISCUSIÓN

En esta sección se discuten los resultados.

Como se aprecia en la Figura 1, la mínima distancia teórica alcanzable cuando trabajamos en la banda de 125 KHz es de 3489 m., usando el SF7 y una imagen de tamaño CGA, y la máxima 11077 m., usando un SF12 y una imagen QXGA. Según la Tabla III, se puede apreciar que cada vez que aumentamos el SF en 1, aumentamos el ToA para poder realizar la transmisión entre un 175-185%. Es decir, cada vez que necesitemos alcanzar una mayor distancia de transmisión, casi duplicamos el ToA. Según esto, en áreas donde ubiquemos nuestros nodos y sean susceptibles de que se produzcan colisiones, cada vez que queremos ampliar nuestra cobertura en 1 o 2 Km., veremos mermada la capacidad de transmisión casi a la mitad, según se empleen los diferentes SF. También se puede apreciar en la Tabla III que el consumo de energía necesario para transmitir con cada uno de los SF sigue la misma regla que los de ToA. Es decir, a mayor ToA mayor necesidad de consumo de energía. Podemos observar que, la mínima cantidad de energía necesaria para enviar una imagen en blanco y negro de tamaño CGA con SF7 es de 1122 mJ, mientras que, si la queremos enviar en color el consumo de energía aumenta a 8959 mJ. De la misma forma, la mínima cantidad de energía necesaria para enviar una imagen en blanco y negro de tamaño QXGA con SF7 es de 55034 mJ, mientras que, si la queremos enviar en color el consumo de energía aumenta a 1320806 mJ.

Se producen resultados similares al emplear la banda de 250 KHz. En este caso, la distancia mínima alcanza es de 2769 m., usando el SF7, y la máxima 8792 m., usando un SF12. Respecto al ToA y el consumo de energía, los resultados son similares a nivel proporcional a los presentados al transmitir en la banda de 125 KHz.

Los resultados observados en la banda de 500 KHz., son similares a los anteriores. La diferencia más apreciable está relacionada con la distancia de transmisión. En este caso, la distancia mínima alcanzable es de 2189 m., usando el SF7, y la máxima 6978 m., usando un SF12. Respecto al ToA y el consumo de energía, los resultados son similares a nivel proporcional a los presentados al transmitir en las bandas de 125 y 250 KHz.

En relación, al número de paquetes enviados, es muy importante considerar la necesidad de que las imágenes se envíen en blanco y negro o color, pues aplicaremos un coeficiente multiplicador de 24 en el tamaño del archivo generado, según se puede apreciar en la Tabla II. Además, en el entorno real, se deberá tener en cuenta la capacidad de procesamiento del nodo, pues se podrían aplicar técnicas de compresión del archivo, donde se puede llegar a comprimir con JPEG, logrando una compresión de 10: 1 con poca pérdida perceptible en la calidad de la imagen [13].

V. CONCLUSIONES

LoRa es una tecnología de largo alcance que está creciendo en popularidad cada vez más. Asimismo, su uso se extiende a diversos ámbitos como la agricultura de precisión. Sin embargo, la transmisión de imágenes con LoRa apenas se ha estudiado. En este estudio, hemos evaluado la distancia máxima que se puede alcanzar para diversos anchos de banda y SF en la frecuencia de 868 MHz. Asimismo, se ha determinado el número de paquetes necesarios para enviar distintos formatos de imagen. El tiempo necesario para enviar cada una de las imágenes se ha evaluado también. Finalmente, se ha proporcionado la energía necesaria para transmitir completamente cada tipo de imagen. Los resultados muestran que un ancho de banda mayor está relacionado con una menor distancia y un tiempo de transmisión y consumo energético menor. A su vez, un SF mayor implica un mayor ToA y un mayor consumo energético.

Como trabajos futuros, se realizarán pruebas prácticas en las que se evalúe el rendimiento de la transmisión e imágenes con LoRa para incluir la funcionalidad de la transmisión de imágenes en aplicaciones de agricultura de precisión y monitorización de bosques como las propuestas en [14] y [15].

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por la Unión Europea a través del proyecto ERANETMED (Euromediterranean Cooperation through ERANET joint activities and beyond) ERANETMED3-227 SMARTWATIR y a través del Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema de I+D+i y del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020 (Cod. Proyecto: PID2020-114467RR-C33). Este trabajo también ha sido parcialmente financiado por la Universitat Politècnica de València a través del programa post-doctoral PAID-10-20.

REFERENCIAS

[1] LoRa Documentation. Available online: https://lora.readthedocs.io/en/latest/ (accessed on 24/06/2021).

[2] Benites, B.; Chávez, E.; Medina, J.; Vidal, R.; Chauca, M. LoRaWAN applied in Swarm Drones: A focus on the use of fog for the management of water resources in Lima-Peru. In Proceedings of the 5th International Conference on Mechatronics and Robotics Engineering, Rome, Italy, 16–19 February 2019; pp. 171–176., 21.

[3] Sanchez-Iborra, R.; Sanchez-Gomez, J.; Ballesta-Viñas, J.; Cano, M.D.; Skarmeta, A.F. Performance evaluation of LoRa considering scenario conditions. Sensors 2018, 18, 772.

[4] Ochoa, M.N.; Guizar, A.; Maman, M.; Duda, A. Evaluating LoRa energy efficiency for adaptive networks:From star to mesh topologies. In Proceedings of the IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Rome, Italy, 9–11 October 2017. DOI: 10.1109/WiMOB.2017.8115793

[5] Semtech Application Note AN1200.22. 2015. LoRa Modulation Basics. Available online: http://wiki.lahoud. fr/lib/exe/fetch.php?media=an1200.22.pdf (Last accessed on 24/06/2021)

[6] Semtech Application Note AN1200.13. 2013. LoRa Modem Designer's Guide. Available online: https://www.rs-online.com/designspark/rel-assets/ds-assets/uploads/knowledge-items/application- notes-for-the-internet-of-things/LoRa%20Design%20Guide.pdf (accessed on 24/06/2021)

[7] Waret, A.; Kaneko, M.; Guitton, A.; El Rachkidy, N. LoRa throughput analysis with imperfect spreading factor orthogonality. IEEE Wirel. Commun. Lett. 2018, 8, 408–411.

[8] Croce, D.; Gucciardo, M.; Tinnirello, I.; Garlisi, D.; Mangione, S. Impact of spreading factor imperfect orthogonality in lora communications. InInternational Tyrrhenian Workshop on Digital Communication; Springer: Cham, Switzerland, 2017; pp. 165–179.

[9] Croce, D.; Gucciardo, M.; Mangione, S.; Santaromita, G.; Tinnirello, I. Impact of LoRa ImperfectOrthogonality: Analysis of Link-Level Performance. IEEE Commun. Lett. 2018, 22, 796–799.

[10] Mikhaylov, K.; Petäjäjärvi, J.; Janhunen, J. On LoRaWAN scalability: Empirical evaluation of susceptibility tointer-network interference. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017.

[11] Pham, C. Building Low-Cost Gateways and Devices for Open LoRa IoT Test-Beds. In Proceedings of the International Conference on Testbeds and Research Infrastructures, Hangzhou, China, 14–15 June 2016; pp. 70–80; DOI: 10.1007/978-3-319-49580-4_7

[12] Bouguera, T.; Diouris, J.; Chaillout, J.; Jaouadi, R.; Andrieux, G. Energy Consumption Model for Sensor Nodes Based on LoRa and LoRaWAN. Sensors 2018, 18, 2104.

[13] J. M. Jimenez, L. Parra, L. García, J. Lloret, P. V. Mauri and P. Lorenz, "New Protocol and Architecture for a Wastewater Treatment System intended for Irrigation", Applied Sciences, Vol. 11, No. 8, pp. 3648, 2021.

[5] Semtech Application Note AN1200.22. 2015. LoRa Modulation Basics. Available online: http://wiki.lahoud. fr/lib/exe/fetch.php?media=an1200.22.pdf (Last accessed on 24/06/2021)

[6] Semtech Application Note AN1200.13. 2013. LoRa Modem Designer's Guide. Available online: https://www.rs-online.com/designspark/rel-assets/ds-assets/uploads/knowledge-items/application- notes-for-the-internet-of-things/LoRa%20Design%20Guide.pdf (accessed on 24/06/2021)

[7] Waret, A.; Kaneko, M.; Guitton, A.; El Rachkidy, N. LoRa throughput analysis with imperfect spreading factor orthogonality. IEEE Wirel. Commun. Lett. 2018, 8, 408–411.

[8] Croce, D.; Gucciardo, M.; Tinnirello, I.; Garlisi, D.; Mangione, S. Impact of spreading factor imperfect orthogonality in lora communications. InInternational Tyrrhenian Workshop on Digital Communication; Springer: Cham, Switzerland, 2017; pp. 165–179.

[9] Croce, D.; Gucciardo, M.; Mangione, S.; Santaromita, G.; Tinnirello, I. Impact of LoRa ImperfectOrthogonality: Analysis of Link-Level Performance. IEEE Commun. Lett. 2018, 22, 796–799.

[10] Mikhaylov, K.; Petäjäjärvi, J.; Janhunen, J. On LoRaWAN scalability: Empirical evaluation of susceptibility tointer-network interference. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017.

[11] Pham, C. Building Low-Cost Gateways and Devices for Open LoRa IoT Test-Beds. In Proceedings of the International Conference on Testbeds and Research Infrastructures, Hangzhou, China, 14–15 June 2016; pp. 70–80; DOI: 10.1007/978-3-319-49580-4_7

[12] Bouguera, T.; Diouris, J.; Chaillout, J.; Jaouadi, R.; Andrieux, G. Energy Consumption Model for Sensor Nodes Based on LoRa and LoRaWAN. Sensors 2018, 18, 2104.

[13] Haines, Richard F.; Chuang, Sherry L. The effects of video compression on acceptability of images for monitoring life sciences experiments (Technical report). NASA. NASA-TP-3239, A-92040, NAS 1.60:3239. 1 July 1992.

[14] J. M. Jimenez, L. Parra, L. García, J. Lloret, P. V. Mauri and P. Lorenz, "New Protocol and Architecture for a Wastewater Treatment System intended for Irrigation", Applied Sciences, Vol. 11, No. 8, pp. 3648, 2021.

[15] S. Sendra, L. García, J. Lloret, I. Bosch and R. Vega-Rodríguez, "LoRaWAN network for fire monitoring in rural environments", Electronics, Vol. 9, No. 3, pp. 531, 2020.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Mercado de datos IoT sustentado en tecnologías Blockchain

Jorge Lanza, Iván González, Luis Sánchez, Juan Ramón Santana, Pablo Sotres
Departamento Ingeniería de Comunicaciones
Universidad de Cantabria
Avda. Los Castros S/N - Santander - 39005 Cantabria.
{jlanza, igonzalez, lsanchez, jrsantana, psotres}@tlmat.unican.es

El despliegue de infraestructuras de la Internet de las Cosas ha supuesto una revolución en la adquisición de información del contexto alrededor de los servicios provistos para y por los usuarios. Sin embargo, las soluciones globales más comunes se basan en metodologías propietarias que no implementan mecanismos que acrediten y garanticen el origen y futuro uso confiable de los datos generados y compartidos. Estas, además, excluyen al usuario como fuente de los datos de contexto de la cadena de valor. Este artículo describe y evalúa un ecosistema de gestión de datos IoT basado en Blockchain, que garantiza al proveedor el control sobre quién, cómo y cuándo hace uso de los datos, al tiempo que permite explotar nuevos modelos de negocio y monetización de los mismos.

***Palabras Clave*- IoT, Blockchain, prosumer, mercado**

I. INTRODUCCIÓN

El panorama de soluciones de la Internet de las Cosas (IoT, *Internet of Things*), además de altamente fragmentado, está dominado por soluciones verticales propietarias. Las soluciones estándar se restringen al entorno de la investigación y la experimentación [1-3]. Resultado de los escasos esfuerzos de estandarización, los usuarios, emprendedores y PyME se encuentran inmersos en un monopolio comercial que reduce sus expectativas a la hora de afrontar soluciones innovadoras en la plétora de potenciales escenarios ante la dificultad de poder aplicar economías de escala.

Las tradicionales infraestructuras IoT exportan información que, en la mayoría de los casos, no es considerada sensible. Sin embargo, el acercamiento de estas redes al entorno de la empresa o del usuario requieren de políticas claras y robustas en términos de privacidad y protección de datos que garanticen la confianza, extendiendo los actuales paradigmas centralizados basados en entidades de confianza hacia soluciones descentralizadas, federadas y transversales.

Lograr superar estas barreras permitirá, entre otras cosas, ampliar el abanico de actores interesados en el despliegue de soluciones de valor añadido en el ámbito de la IoT. De hecho, la colaboración entre ellos, bajo un modelo de co--creación, podría suponer la creación de un nuevo ecosistema, basado en la confianza y la diversidad, donde se fomente y sea habitual la creación de nuevas aplicaciones disruptivas. Monetizar o incentivar económicamente tanto la cesión de los datos como el uso de los servicios jugará un papel determinante en la sostenibilidad, y, por tanto, en el éxito de este nuevo modelo.

La tecnología Blockchain habilita los mecanismos para desplegar soluciones totalmente descentralizadas que proporcionen trazabilidad garantizada del ciclo de vida de servicios y los datos subyacentes, es decir, que se habilitan los mecanismos para gestionar de forma confiable las transacciones de datos y/o monetarios entre los distintos actores implicados. Por tanto, Blockchain puede suponer una respuesta adecuada a los requerimientos de calidad de la información, de confianza en la fuente de los datos y control del potencial uso de éstos, etc. Todo ello permitirá hacer del ecosistema anteriormente descrito una realidad.

No obstante, ha de evitarse la generación de infraestructuras Blockchain independientes y de carácter vertical que harían emerger nuevamente las problemáticas anteriormente señaladas. Es por esto que exportar la solución como un servicio, bajo el concepto *as a Service*, permitiría integrar cualquier ecosistema existente con necesidades confianza distribuida, de forma rápida y sencilla.

Este artículo presenta una plataforma, basada en Blockchain, que exporta un mercado de datos IoT, el cual habilita intercambios transparentes, seguros y confiables entre productores y consumidores de datos. Este mercado (BIDM, *Blockchain-based IoT Data Marketplace*) genera un ecosistema a través del que no solo se implementa el tradicional contrato de compraventa que establece una compensación (precio) por un servicio o bien, sino que lo extiende para obligar a ambas partes a cumplir unas

condiciones y requerimientos previos y a futuro, desde el punto de vista de reputación, usos del bien adquirido, compensaciones en caso de incumplimiento, etc.

En este sentido, el artículo describe la arquitectura funcional de la plataforma que sustenta el BIDM y los procedimientos para la provisión y consumo de información según las premisas anteriormente expuestas. Además, se incluye la implementación y despliegue de una instancia totalmente funcional del BIDM integrada en el ámbito de una plataforma IoT a gran escala [4], de forma que se pueda evaluar su operativa en condiciones reales.

II. Estado del arte

El acceso a los flujos de datos generados por las infraestructuras IoT puede circunscribirse dentro de los modelos de computación en la nube que consideran que cualquier información está disponible remotamente a través de tecnologías web [5], y más específicamente al modelo sensado como Servicio (SaaS, *Sensing as a Service*) mediante el cual las aplicaciones adquieren la información de contexto necesaria usando estas arquitecturas orientadas al servicio.

Centrado en el ecosistema de las ciudades inteligentes, Diaz et al. [6] plantean una arquitectura funcional para estos escenarios sustentada en tres actores: generadores de datos, proveedores de servicios y consumidores de datos y servicios. El concepto de BIDM se puede extrapolar a este planteamiento puesto que considera el comercio regulado de datos entre los dueños de la infraestructura IoT, aquellos que exponen la información agregada de contexto haciendo uso de servicios, y los que hacen uso inteligente de la información disponible.

Originalmente, la información en crudo, sin procesar, proveniente directamente de los sensores era accesible a través de sistemas centralizados en la nube. Este modelo de mercado de datos, que podría llegar a despertar ciertas reticencias en términos de privacidad para productores de datos que no fueran, a su vez, los propios gestores de esos sistemas centralizados, evolucionó hacia soluciones Peer-to-Peer (P2P) [7], cuyo carácter distribuido minimiza la probabilidad de fallo total del sistema al evitar el potencial único punto de error de las anteriores soluciones. Sin embargo, aunque distribuidas, no daban respuesta a los problemas de escalabilidad que supone proporcionar un registro globalmente compartido e interoperable entre plataformas IoT, aspecto este considerado por la propuesta BIDM presentada en este artículo.

Adicionalmente, otro aspecto a considerar en el despliegue de mercados de datos distribuidos es la necesidad de disponer de un modelo de gestión de la confianza, de forma que las transacciones sean validadas sin necesidad de una entidad de confianza central. Yan et al. [8] y Perera et al. [9] trazan las métricas y propiedades (seguridad, confiabilidad, disponibilidad, precisión, etc.) que los metadatos asociados a la información de contexto generada en el ámbito de la IoT deben incluir en aras de proporcionar confianza necesaria para su uso. Éstas pueden ser fácilmente añadidas al modelo de datos que se considera en el BIDM. Es más, el BIDM amplía la confianza más allá del dato, considerando también a los propios productores y consumidores.

Enfocándose en soluciones más recientes en las que el soporte distribuido se sustenta en tecnologías Blockchain, se observa un creciente interés en las soluciones que exploran la integración de las mismas con la IoT [10][11]. La mayoría de ellas se centran en verticales específicos (vehículos autónomos, salud, trazabilidad alimentaria, etc.) [12]-[15], en lugar de buscar una solución transversal como la que se trata en este artículo.

No obstante, existen diversas arquitecturas para dar soporte a mercados de datos con enfoques similares al descrito aquí. La plataforma MARSA [16], si bien comparte objetivo en tanto en cuanto categoriza los productores y sus datos vinculados y establece los mecanismos para comerciar con ellos, considera el uso de una entidad central que actúa como intermediario de confianza entre las partes involucradas en una transacción. La propuesta descrita por Misura [17] también apuesta por un agente que monitoriza las transacciones, pero en este caso, el intercambio es directo entre las partes. Por tanto, se trata de una solución híbrida en línea con el BIDM, aunque hay que destacar que en el caso del BIDM el agente central actúa únicamente en el almacenaje de los datos, basando el resto de las transacciones en un sistema distribuido sobre Blockchain. Adicionalmente, el BIDM da soporte para la transmisión continua de datos a diferencia de la solución de Misura basada en el modelo petición-respuesta.

Por último, destacar el trabajo de Ozyilmaz et al. [18] que describe un mercado como un conjunto de contratos inteligentes (*smart-contract*) ejecutándose en una red Blockchain Ethereum, que acceden a los datos almacenados en Swarm [19], un entorno distribuido de almacenamiento de información. Si bien la solución presenta grandes similitudes con el BIDM, es en el almacenamiento donde difiere. El BIDM, en su objetivo de lograr la interoperabilidad y la federación entre entornos IoT, se apoya en una solución estándar y de código abierto ampliamente adoptada por la comunidad IoT para el desarrollo de ecosistemas inteligentes como es FIWARE [20].

III. Arquitectura

La plataforma que se presenta en este artículo tiene como uno de sus objetivos principales la integración de la tecnología Blockchain en el ámbito de cualquier ecosistema IoT actual o futuro, y ofrecer a través de ella los mecanismos para gestionar el comercio de datos de forma segura y confiable. Se explotan las propiedades inherentes de Blockchain para permitir la trazabilidad de los datos y las operaciones, logrando así satisfacer las siguientes condiciones fundamentales de diseño:

- Las medidas o datos son generados y exportados por entidades de confianza.
- El pago por el acceso a la información se realiza en un momento específico.
- Los compradores reciben la información adquirida.

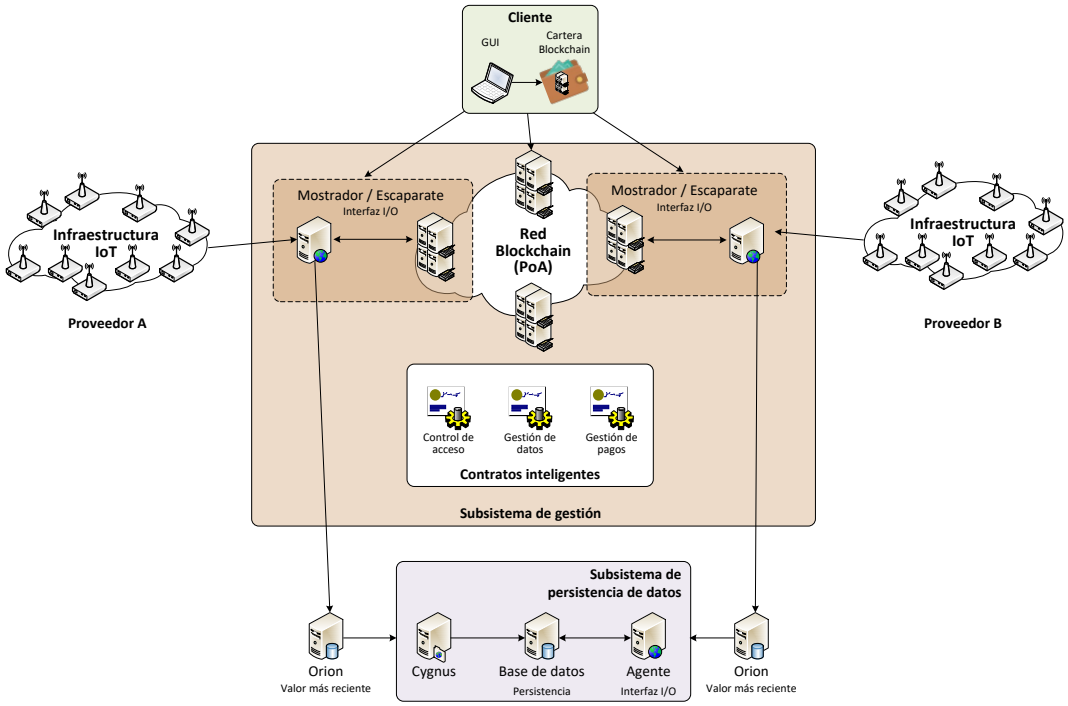


Fig. 1. Arquitectura funcional de la solución

- Sólo aquellas entidades y/o usuarios autorizados podrán acceder de forma segura a la información.

Todo ello redunda en una total transparencia en la operativa del sistema, tanto para los consumidores o clientes como para los productores o infraestructuras IoT.

A. *Arquitectura funcional*

La Fig. 1 muestra la arquitectura funcional del BIDM y las relaciones entre los diferentes elementos que la conforman. Se identifican varios grupos funcionales: por un lado, el subsistema gestor de la plataforma y el de persistencia de datos y, por otro, las entidades externas que estarán vinculadas a los proveedores o a los consumidores de datos. Se ha optado por incluir los elementos que conforman la red Blockchain entre los componentes del subsistema gestor ya que toda la inteligencia de la operativa del mercado tiene como núcleo central esta tecnología. La red Blockchain sobre la que se cimenta el BIDM, ya sea pública o privada, se basa en el protocolo de consenso de Prueba de Autoridad (PoA) y admite la definición de contratos inteligentes. Esta dupla reduce las posibles implementaciones a redes operando con tecnología basada en Ethereum, ya sea la propia Ethereum u otras como Quorum, etc. PoA hace uso de la identidad y la reputación de los validadores como garantía de velar por el buen funcionamiento, la transparencia y confiabilidad de las operaciones dentro de la red. Considerando el objetivo de sustentar el mercado de datos dentro de una federación de infraestructuras IoT, se estima que la opción natural para dar soporte al BIDM es el empleo de este método de consenso unido al despliegue de una red Blockchain permisionada, donde todos los entornos IoT federados tienen el mismo peso y, por tanto, el mismo grado de confianza. Adicionalmente, PoA, frente a otros protocolos de consenso como Prueba de Trabajo (PoW, *Proof-of-Work*), se adapta mejor a la naturaleza asíncrona de la IoT puesto que permite fijar la periodicidad de los bloques minados en función de las tasas máximas y mínimas de publicación de datos, minimizando así el número de bloques vacíos almacenados en la cadena de bloques.

Profundizando en los componentes funcionales principales, encontramos por un lado el mostrador o escaparate como punto de interconexión con el exterior y el subsistema de persistencia de datos:

- Mostrador o escaparate, considerado como un interfaz de I/O, es el elemento a través del cual se recolectan las medidas remitidas por los productores de datos, se publicitan a los potenciales consumidores y, finalmente, las sirve a solicitud de estos últimos. Asimismo, redirige la información al subsistema de persistencia, y la indexa y referencia dentro de la cadena de bloques para facilitar su posterior búsqueda y acceso. Este componente se puede dividir en un proxy de entrada o API que procesa la información recibida y un nodo Blockchain como punto de entrada a red y elemento habilitador de la interacción con los contratos inteligentes que se ejecutan en ella. Podrán existir instancias por cada proveedor y actuarán como cartera de los mismos. La disponibilidad de diferentes puntos de entrada a la red Blockchain reduce los potenciales problemas de escalabilidad. La interfaz visual puede ser adaptada en cada instancia, si bien habilitará el acceso a la información de forma global.
- El Subsistema de persistencia de datos es el almacén de la información, sustentado en habilitadores genéricos del ecosistema FIWARE como Orion y Cygnus. Incluye también una base de

datos indexada por el hash del bloque y accesible a través de un agente web. El uso de los modelos de datos estándar e interoperables permitirá compartir infraestructura entre diferentes proveedores o federar las propias.

B. Procedimientos

Además de estos elementos, se despliegan tres contratos inteligentes que son los que aportan la confianza en el BIDM durante las operaciones de control de acceso, almacenamiento de datos y pago por uso. El primero de ellos, limita el registro de medidas o datos en la plataforma a sensores o productores de confianza, del mismo modo que restringe el acceso a la información a clientes debidamente registrados y validados. El registro de estos usuarios, tanto productores como consumidores, se hace mediante un procedimiento de gestión de usuarios cuyas credenciales, vinculadas a sus cuentas Ethereum y/o certificados digitales, se integran con la base de datos de usuarios a la que accede el contrato de control de acceso.

El contrato vinculado al almacenamiento de datos gestiona cómo se guarda la información en la cadena de bloques. El modelo de datos empleado incluye, además del hash de la medida, como parámetro de indexación y garantía de integridad, una referencia absoluta a su localización en el subsistema de persistencia de datos, de forma que se pueda recuperar de forma total o parcial según lo acordado en la transacción de compra. Además, se incorporan una serie de metadatos descriptivos de la medida, que ayudarán a publicitarla en el propio mercado. En la implementación que se ha realizado, la información contenida en estos metadatos se refiere al fenómeno físico medido en la observación, al sensor del que proviene y al precio asignado a la medida, pero desde el punto de vista de diseño, se podría extender para incluir características de calidad de la medida, reputación del productor, etc. Por último, el contrato de pagos gestiona la transacción de adquisición de una medida, evitando pagos duplicados y garantizando la recepción de la misma según las condiciones de compra.

A continuación, se detalla la interacción entre ellos para dar respuesta a las necesidades de autenticación y acceso a la información.

Para garantizar el origen de la información almacenada, únicamente sensores o proveedores de confianza previamente autorizados pueden interactuar con la plataforma. El proceso de autorización consiste en asociar un identificador único a estas fuentes de información. El proceso de asociación puede realizarse de múltiples formas, entre ellas delegando la confianza en el propio mercado de datos. En el caso que se describe en este artículo, se ha optado por generar las credenciales fuera de línea y grabarlas de forma segura en los dispositivos, es decir, el gestor particular del BIDM genera un identificador y una contraseña única para cada dispositivo. Adicionalmente se provee la clave pública o el certificado digital del BIDM.

Las solicitudes de registro de medidas enviadas por los proveedores deben, por tanto, incluir, además de la propia medida en formato NGSIv2 [21], las credenciales que

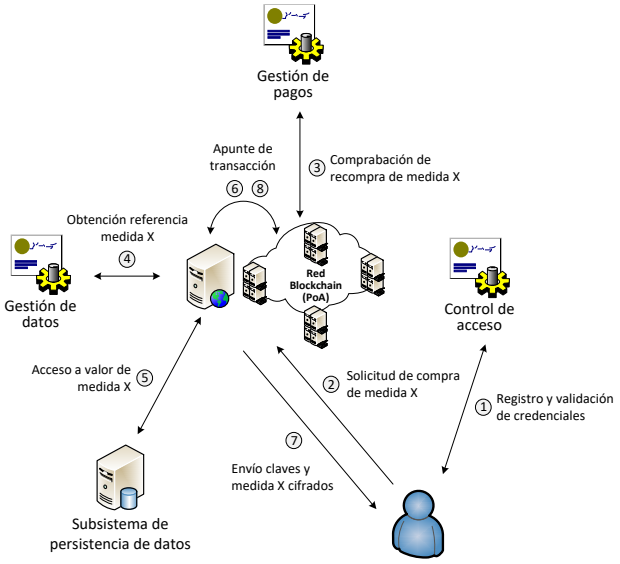


Fig. 2. Proceso de acceso a medidas

permitan validar la identidad del proveedor. Para evitar comprometer la confidencialidad de las mismas, tanto el identificador como la contraseña se cifran con la clave pública del BIDM.

A la recepción de la medida, y una vez comprobado que proviene de un productor autorizado, se comprueba la integridad y validez de la información recibida y se adapta el documento eliminando información no relevante o redundante. Posteriormente, se calcula el hash del documento final y redirige para su almacenamiento.

Una vez confirmado el almacenamiento de la medida por parte del subsistema de persistencia de datos, que incluye un enlace único y permanente al documento almacenado, se procede a inyectar el hash, la referencia y el conjunto de metadatos (descripción, etc.) en la cadena de bloques.

La información residente en la cadena de bloques permite facilitar la búsqueda de datos apoyándose únicamente en los datos disponibles en ella, al tiempo que se reduce el tamaño de la mismo al externalizar el almacenamiento de los datos de la medida. No obstante, se mantienen las garantías de integridad al incorporar de forma disjunta la referencia a la medida y el resumen de la misma. La confianza en el conjunto de datos que conforman el documento la otorga la propia plataforma y operativa Blockchain.

El proceso de búsqueda se lleva a cabo por parte de los potenciales clientes. Los clientes son entidades también registradas y validadas en la red Blockchain, pero a diferencia de los proveedores IoT, éstos tienen permisos restringidos a operaciones de lectura. La búsqueda se realiza directamente sobre la propia cadena de bloques, si bien, como alternativa, se puede considerar la búsqueda directa en el entorno de persistencia a través de un API generada para tal efecto. El resultado de ambas búsquedas será el hash del documento que permita referenciar a la cadena de bloques.

Identificada la medida o conjunto de medidas que se quieren adquirir, se procede a continuación a la compra, la cual otorga el derecho de acceso a los valores de las

mismas. La Fig. 2 muestra el desarrollo del proceso que deben seguir los consumidores, en el que están involucrados todos los contratos inteligentes.

El cliente comienza el proceso de compra iniciando una transacción de compra con el contrato inteligente de pagos (paso 2 en la Fig. 2), quien apoyado en el contrato de control de acceso garantiza que se trata de un cliente autorizado. Entre los datos incluidos en la transacción de compra se encuentra además de los identificadores de cliente, el identificador único o hash de las medidas que se desean adquirir. A partir de éste, puesto que está almacenado también en la cadena de bloques, el contrato obtiene la referencia única (i.e. URL) a la medida. Teniendo en cuenta que el agente del subsistema de persistencia ofrece una interfaz para acceso a la información basada en servicios RESTful, tras la correspondiente petición HTTP se obtiene el valor de la medida.

La última fase del proceso garantiza que el proceso de pago finaliza únicamente cuando el comprador accede a los datos. Para ello, antes de proceder a su entrega se acondicionan los datos para garantizar que únicamente el adquiriente puede acceder a ellos.

De este modo, la medida se cifra empleando un algoritmo de cifrado simétrico (i.e. AES) utilizando una clave suficientemente robusta generada de manera aleatoria. Esta clave se facilita al cliente cifrándola con la clave pública incluida en su propio certificado. De esta forma se garantiza que únicamente dicho cliente es capaz de descifrar la clave y, por tanto, la medida.

Adicionalmente, con objeto de que el administrador del BIDM pueda acceder a la medida en caso de disputa, también se cifra la clave con la clave pública vinculada al BIDM. De esta forma, si el cliente reporta que la medida no se ajusta a lo adquirido, el administrador podrá comprobar qué valores se le remitieron.

Toda esta información, claves y medida cifrada, se incluye en una transacción entre el BIDM y el cliente, que se anota en la red Blockchain. Dada la naturaleza pública de la información incluida en los bloques de la Blockchain, cualquier usuario con credenciales autorizadas tiene acceso a esta información vinculada a la compra. Sin embargo, únicamente el cliente que ha realizado la compra podrá acceder a la medida en sí ya que esta se almacena en el subsistema de persistencia de la información protegida según el procedimiento anterior.

En este momento, se da por concluida la transacción de pago reflejándolo con el apunte correspondiente en la cadena de bloques.

El proceso descrito implica el almacenaje de la medida en la cadena de bloques de forma cifrada. Inicialmente puede resultar incongruente con la premisa inicial que imponía almacenar las medidas fuera de la cadena de bloques en el componente de persistencia de datos. Sin embargo, el volumen de datos generados por los proveedores excede con mucho las compras que se realicen. Es por ello que se ha optado por esta metodología pues se considera que los beneficios en cuanto a garantía de disponibilidad compensan las necesidades adicionales de espacio de almacenaje.

Como proceso adicional, el cliente puede comprobar que la medida obtenida coincide con la solicitada gracias al resumen que se incluye de ésta en la Blockchain en el momento de publicar la medida.

Para concluir, señalar que los procesos descritos permiten la trazabilidad completa del ciclo de vida de un dato generado por cualquier proveedor IoT, tanto en el propio proceso de generación como en tantos procesos de adquisición como consumidores haya interesados en dicha observación.

IV. Prueba de concepto

A. *Entorno de desarrollo y despliegue*

Para la validación de la solución descrita en este artículo, se han desarrollado e integrado todos los componentes funcionales del BIDM en una implementación de prueba de concepto. El despliegue se ha realizado en un entorno virtual empleando contenedores Docker para facilitar la replicabilidad del sistema.

Se ha tomado la decisión de emplear una red Blockchain privada basada en Ethereum Clique [22], configurada para usar el protocolo de consenso PoA disponible. La red desplegada para la validación consta de cuatro nodos, dos de ellos actuando como validadores y los otros dos como puntos de entrada a la red (sin permisos de validación) para un proveedor, y una cartera de consumidor de información de contexto respectivamente. Adicionalmente se despliega un quinto nodo como nodo inicializador de apoyo al descubrimiento de la configuración de la red distribuida. Si bien puede no considerarse indispensable, la presencia de este nodo reduce y facilita significativamente el proceso de descubrimiento de nodos en redes de gran tamaño, motivo por el cual se incluye como apoyo al soporte a la escalabilidad.

Los contratos inteligentes se han desarrollado en Solidity y desplegado directamente en la red Blockchain, estando por tanto disponibles desde el inicio de la misma.

Finalmente, para el almacenamiento persistente de la información de contexto recolectada de las infraestructuras IoT se han empleado los habilitadores de FIWARE Orion Context Broker y Cygnus, ambos desplegados mediante sendos contenedores Docker.

La interfaz del mostrador o escaparate y la cartera de los clientes se ha realizado empleando tecnologías web, la lógica del servicio web en Nodejs y el interfaz de usuario en HTML y Javascript.

B. *Integración y validación en infraestructura IoT real*

La validación de la implementación realizada se enmarca dentro de la infraestructura IoT disponible en la ciudad de Santander (España), gestionada en el ámbito del proyecto SmartSantander.

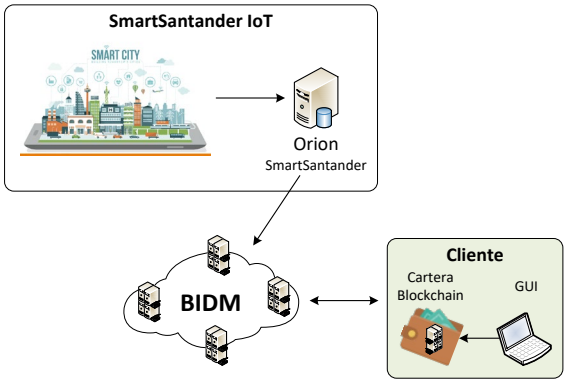


Fig. 3. Integración del BIDM en la infraestructura de SmartSantander

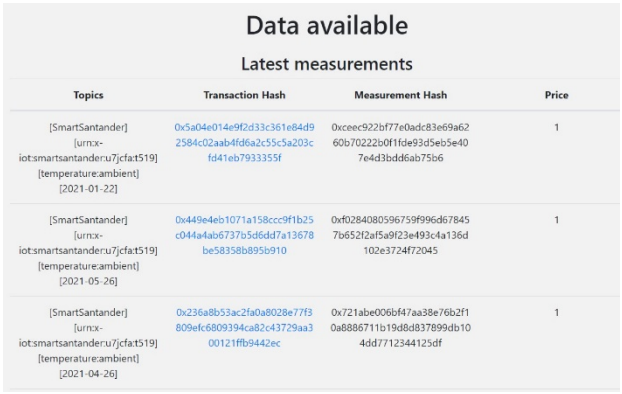


Fig. 4. Datos disponibles en la cadena de bloques

Tal como se muestra en la Fig. 3, el BIDM, para el que únicamente se ha desplegado una instancia del escaparate de datos, se suscribe a las medidas generadas por los sensores disponibles en la ciudad de Santander y exportadas a través de un Orion Context Broker dispuesto a tal efecto. Por tanto, la infraestructura de SmartSantander actúa como proveedor de confianza autorizado en el BIDM. La información recolectada por el BIDM, en formato NGSIv2, es tratada y almacenada tanto en la cadena de bloques asociada al BIDM como en el entorno de persistencia según se ha descrito, para que cualquier cliente pueda solicitar su adquisición. En la Fig. 4, se muestra un ejemplo de cómo se almacenan las medidas en la cadena de bloques. A través de la lectura de la cadena de bloques, el consumidor de información de contexto, además de al hash, tiene acceso a la información descriptiva de dicha medida (en el caso de la implementación realizada, esta información era el identificador del sensor que generó la medida, una marca temporal y el precio fijado para su adquisición.
Un cliente interesado y autorizado puede realizar la compra a través de su cartera virtual.

La Fig. 5 muestra el interfaz de una sencilla cartera, implementada como una página web desde la cual se puede explorar la Blockchain y ver la información incluida en el Mostrador/Escaparate (i.e. la información incluida en los bloques de publicación de medidas). A través de ella, era posible obtener el resumen de la medida deseada y, con él, lanzar una petición de compra, resultado de la cual obtendrá la medida cifrada y la correspondiente clave para descifrarla. Lo que en esta prueba de concepto se implementa como una web desde la cual se seleccionan medidas y se ejecutan compras, en un caso de uso característico podría ser una aplicación móvil que ofrezca un servicio de guía turística en la ciudad y entre sus servicios esté el de recomendación de rutas para ir de un lugar a otro de la misma. La aplicación actuará como cliente y dispondrá de su propia cartera virtual. Cuando el usuario solicite una recomendación para un desplazamiento, la aplicación comprará en el BIDM la información relativa a los tiempos de llegada de los autobuses urbanos a la parada más próxima, las bicicletas disponibles en las estaciones del servicio público de alquiler más cercanas al origen y destino, y el número de plazas de aparcamiento en las inmediaciones del destino, para con ello ofrecer diferentes alternativas de viaje al usuario. De otra parte, el BIDM publica y permite el acceso a la información toda vez se ha cumplimentado el pago por la misma y se satisfacen las condiciones de uso impuestas. El BIDM garantiza, en cierto modo, la calidad de la información y su procedencia. En esta situación, todos los participantes en la cadena del servicio obtienen un beneficio dentro de un ecosistema de confianza mutua. La remuneración obtenida por la venta de los datos permitirá al proveedor mejorar y/o ampliar su infraestructura, lo que redunda en el beneficio del desarrollador de la aplicación que podrá ofrecerla en mejores condiciones a un mayor número de usuarios.

La Fig. 6 muestra una captura de la aplicación de cartera desarrollada en la que se muestra la información disponible a la finalización de una compra en la que se incluye el hash de la transacción que incorpora la clave simétrica de cifrado a la cadena de bloques (KeyTxHash) y el hash de la transacción vinculada a la transferencia de la medida solicitada cifrada (DataTxHash), además del propio hash del evento informativo del final de la compra (EventTxHash). Si bien mediante estas referencias el usuario podría acceder al valor de la medida, la aplicación, con el objetivo de facilitar la operativa al cliente, muestra directamente la medida en claro. En el ejemplo, se proporciona una medida de un sensor que monitoriza el flujo de tráfico en una posición específica.

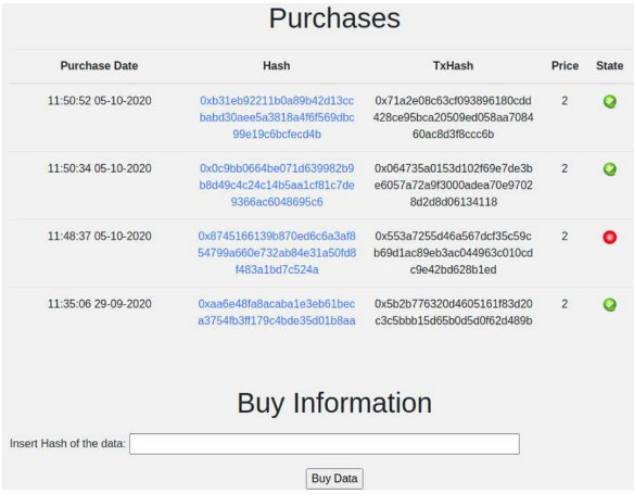


Fig. 5. Interfaz de usuario de la cartera de usuario

Fig. 6. Ejemplo de medida adquirida por un cliente

V. Conclusiones

Este articulo presenta una plataforma que habilita un mercado de datos IoT descentralizado mediante el uso de tecnología Blockchain. A través del BIDM las infraestructuras IoT y las aplicaciones consumidoras de datos de contexto pueden intercambiar información de forma confiable y transparente.

La solución propuesta y descrita en el artículo combina ecosistemas IoT basados en los bloques definidos en el ámbito de Conecting Europe's Facilities (CEF), como es el Orion Context Broker, con la tecnología Blockchain para crear un novedoso entorno que permita la monetización de los flujos de datos con garantías de trazabilidad en todo el ciclo de vida del dato. Gracias a los mecanismos de autenticación y autorización, como a las inherentes propiedades de la tecnología Blockchain se puede garantizar la veracidad, fiabilidad y calidad de los datos.

La solución aborda los potenciales problemas de escalabilidad que podrían resultar del elevado e insostenible crecimiento del tamaño de la cadena bloques. Para ello, se aplica una estrategia de almacenamiento de datos denominada off-chain. En lugar de almacenar todos los datos IoT en la propia Blockchain, estos se almacenan en un entorno especializado e integrado en las infraestructuras IoT, manteniendo en la cadena de bloques únicamente una referencia inmutable a las transacciones realizadas con dichos datos (i.e. registro y compra-venta).

Adicionalmente la solución diseñada da soporte al acceso seguro y garantizado a la información únicamente a aquellos usuarios que, previo pago de la misma, la han adquirido. La combinación de la cadena de bloques, la ejecución de contratos inteligentes en ella y mecanismos de cifrado adicionales lo hacen posible.

Si bien la solución desplegada ha demostrado la viabilidad de un mercado de datos confiable y trazable que combina soluciones estándar de plataformas IoT (i.e. habilitadores funcionales abiertos del ecosistema FIWARE) y las redes Blockchain, capaz de integrarse en un entorno real como es el ecosistema de la ciudad de Santander, aún existen diversas mejoras que aplicar. En este sentido, como trabajo futuro se plantea un exhaustivo análisis de rendimiento y estudio de la escalabilidad del sistema, haciendo hincapié no solo en aspectos vinculados a la seguridad y capacidad de almacenaje de datos, sino también a calidad de servicio y de experiencia de usuario a medida que el volumen de datos aumenta. Además, se plantea la búsqueda de nuevos mecanismos más robustos y amigables de gestionar el ciclo de vida de la información almacenada desde su generación hasta su adquisición y uso por parte del cliente final. Por último, se plantea la extensión del BIDM para dar soporte a consumidores de información con necesidades de tiempo real para lo cual es necesario habilitar un acceso asíncrono (i.e. basado en subscripciones) a las medidas.

Agradecimientos

Este trabajo ha sido realizado en el marco del proyecto FIERCE "Future Internet Enabled Resilient CitiEs" perteneciente al Programa Estatal de I+D+i Orientada a los Retos de la Sociedad (RTI2018-093475-A-I00).

Referencias

[1] H. Hejazi, H. Rajab, T. Cinkler, and L. Lengyel, "Survey of platforms for massive IoT", *2018 IEEE International Conference on Future IoT Technologies, Future IoT 2018*, Mar. 2018, vol. 2018-January, pp. 1–8, doi: 10.1109/FIOT.2018.8325598.

[2] K. J. Singh and D. S. Kapoor, "Create Your Own Internet of Things: A survey of IoT platforms", *IEEE Consumer Electronics Magazine*, vol. 6, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 57–68, Apr. 01, 2017, doi: 10.1109/MCE.2016.2640718.

[3] J. Kim et al., "Standard-based IoT platforms interworking: Implementation, experiences, and lessons learned," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 48–54, Jul. 2016, doi: 10.1109/MCOM.2016.7514163.

[4] L. Sanchez et al., "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, pp. 217–238, 2014.

[5] P. Banerjee et al., "Everything as a service: Powering the new information economy," *Computer*, vol. 44, no. 3, pp. 36–43, Mar. 2011, doi: 10.1109/MC.2011.67.

[6] R. Díaz-Díaz, L. Muñoz, and D. Pérez-González, "Business model analysis of public services operating in the smart city ecosystem: The case of SmartSantander," *Future Generation Computer Systems*, vol. 76, pp. 198–214, Nov. 2017, doi: 10.1016/j.future.2017.01.032.

[7] Z. Huang, X. Su, Y. Zhang, C. Shi, H. Zhang, and L. Xie, "A decentralized solution for IoT data trusted exchange based-on blockchain," in 2017 3rd IEEE International Conference on Computer and Communications, ICCC 2017, Mar. 2018, vol. 2018-January, pp. 1180–1184, doi: 10.1109/CompComm.2017.8322729.

[8] Z. Yan, P. Zhang, and A. Vasilakos, "A survey on trust management for Internet of Things," *Journal of Network and Computer Applications*, vol. 42, pp. 120–134, Jun. 2014, doi: 10.1016/j.jnca.2014.01.014.

[9] C. Perera et al., "Context-aware sensor search, selection and ranking model for internet of things middleware," *2013 IEEE 14th international conference on mobile data management*, vol. 1, pp. 314-322.

[10] S. K. Lo et al., "Analysis of Blockchain Solutions for IoT: A Systematic Literature Review," *IEEE Access*, vol. 7, pp. 58822–58835, 2019, doi: 10.1109/ACCESS.2019.2914675.

[11] A. Panarello, N. Tapas, G. Merlino, F. Longo, and A. Puliafito, "Blockchain and IoT Integration: A Systematic Survey," *Sensors*, vol. 18, no. 8, p. 2575, Aug. 2018, doi: 10.3390/s18082575.

[12] A. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A Decentralized Privacy-Preserving Healthcare Blockchain for IoT," *Sensors*, vol. 19, no. 2, p. 326, Jan. 2019, doi: 475 10.3390/s19020326.

[13] Rathee, Sharma, Iqbal, Aloqaily, Jaglan, and Kumar, "A Blockchain Framework for Securing Connected and Autonomous Vehicles," *Sensors*, vol. 19, no. 14, p. 3165, Jul. 2019, doi: 10.3390/s19143165.

[14] L. Bai, M. Hu, M. Liu, and J. Wang, "BPIIoT: A Light-Weighted Blockchain-Based Platform for Industrial IoT," *IEEE Access*, vol. 7, pp. 58381–58393, 2019, doi: 10.1109/ACCESS.2019.2914223.

[15] Y. P. Tsang, K. L. Choy, C. H. Wu, G. T. S. Ho, and H. Y. Lam, "Blockchain-Driven IoT for Food Traceability with an Integrated Consensus Mechanism," *IEEE Access*, vol. 7, pp. 129000–129017, 2019, doi: 10.1109/ACCESS.2019.2940227.

[16] T. D. Cao, T. V. Pham, Q. H. Vu, H. L. Truong, D. H. Le, and S. Dustdar, "MARSA: A marketplace for real-time human sensing data," *ACM Transactions on Internet Technology*, vol. 16, no. 3, pp. 1–21, May 2016, doi: 10.1145/2883611.

[17] K. Mišura and M. Žagar, "Data marketplace for Internet of Things," *Proceedings of 2016 International Conference on Smart Systems and Technologies*, SST 2016, Dec. 2016, pp. 255–260, doi: 10.1109/SST.2016.7765669.

[18] K. R. Ozyilmaz, M. Dogan, and A. Yurdakul, "IDMoB: IoT data marketplace on blockchain," *Proceedings - 2018 Crypto Valley Conference on Blockchain Technology, CVCBT 2018*, Nov. 2018, pp. 11–19, doi: 10.1109/CVCBT.2018.00007.

[19] V. Tron, "Swarm alpha public pilot and the basics of Swarm," *Ethereum Blog*, 2016. https://blog.ethereum.org/2016/12/15/swarm-alpha-public-pilot-basics-swarm/ (accessed Sep. 21, 2020).

[20] F. Cirillo, G. Solmaz, E. L. Berz, M. Bauer, B. Cheng, and E. Kovacs, "A Standard-Based Open Source IoT Platform: FIWARE," *IEEE Internet of Things Magazine*, vol. 2, no. 3, pp. 12–18, Jan. 2020, doi: 10.1109/iotm.0001.1800022.

[21] FIWARE Data Models, https://www.fiware.org/developers/data-models/

[22] Péter Szilágyi, "Clique proof-of-authority consensus protocol", *Ethereum Improvement Proposal – 225*, March 2017 https://eips.ethereum.org/EIPS/eip-225


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Evolución del Stack IoT: MQTT sobre QUIC

Fátima Fernández[1], Mihail Zverev[1], Pablo Garrido[1], José R. Juárez[1], Josu Bilbao[1], Ramón Agüero[2]
[1]Ikerlan Technology Research Centre, Basque Research Technology Alliance (BRTA), Arrasate/Mondragon, España;
{ffernandez, mzverev, pgarrido, jrjuarez, jbilbao}@ikerlan.es
[2]Dpto. de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, España;
ramon@tlmat.unican.es

En este trabajo se analiza el rendimiento de QUIC como solución de transporte para dar soporte a servicios IoT industriales basados en Message Queuing Telemetry Transport (MQTT). QUIC fue desarrollado por Google para solucionar las limitaciones que presenta el protocolo de transporte dominante, Transmission Control Protocol (TCP). Para estudiar su comportamiento, se han emulado escenarios con características propias a estos entornos industriales, incluyendo QUIC como protocolo de transporte para MQTT y comparándose esta combinación frente a la solución tradicional basada en la pila TCP/TLS/MQTT. Para emular distintas tecnologías y condiciones de red se han empleado contenedores LXC de Linux a modo de dispositivos IoT mediante el simulador de eventos ns-3. Los resultados obtenidos ponen de manifiesto que QUIC podría ser una alternativa interesante como protocolo de transporte en escenarios IoT.

***Palabras Clave*—QUIC; Message Queuing Telemetry Transport (MQTT); Industria 4.0; IoT Industrial (IIoT); Redes Inalámbricas; Entorno de Simulación.**

I. INTRODUCCIÓN

Actualmente, la industria y las empresas están inmersas en la cuarta revolución industrial, denominada Industria 4.0. Este término fue introducido en 2011 por el gobierno alemán, y surgió como una alternativa novedosa hacia la digitalización de la industria en base a cuatro principios: interconexión, transparencia de la información, decisiones descentralizadas y mantenimiento predictivo [1]. Esta revolución trae consigo un aumento de la productividad gracias, entre otras cosas, a la recogida y análisis de datos en tiempo real [2].

Una de las tecnologías clave detrás de esta transformación digital de la industria es el paradigma basado en el Internet de las cosas (*Internet of Things*, IoT), el cual habilita la conexión de dispositivos industriales para recolectar y analizar datos en tiempo real, monitorizar sistemas, intercambiar información, y analizar el entorno industrial [2]. Sin embargo, esta revolución industrial ha desembocado en la necesidad de disponer de nuevos servicios y aplicaciones con estrictos requisitos, como comunicaciones de baja latencia, disponibilidad y fiabilidad en las redes así como reducción en los costes además de ser eficientes energeticamente, canales seguros y la conservación de la privacidad de los datos [3].

Para monitorizar el entorno industrial, se necesita un gran despliegue de sensores y dispositivos conectados entre sí, lo que podría conllevar costes elevados. Además su ubicación en localizaciones diversas y áreas aisladas implica la necesidad de baterías con una vida útil elevada. Las comunicaciones con baja latencia son esenciales para muchas aplicaciones industriales, ya que pueden necesitar de respuestas rápidas para habilitar un proceso de fabricación determinado o garantizar la seguridad de los elementos involucrados. Además, una baja latencia resulta indispensable para habilitar funciones de control remoto en escenarios industriales reales.

MQTT está ganando popularidad en el mundo IoT, y se está convirtiendo en el protocolo de aplicación de facto [4], [5]. Esto se debe principalmente a su fácil integración en los dispositivos y al buen comportamiento que ofrece. Tradicionalmente, funciona sobre el protocolo de transporte TCP [6], que como es sabido ofrece un servicio orientado a la conexión. Sin embargo, TCP no es capaz de adaptarse a la velocidad a la que evoluciona la tecnología, dejando en evidencia la cantidad de desventajas que sufre [7], especialmente las relacionadas con la osificación de los protocolos de internet. Por otro lado, en redes donde la probabilidad de pérdida es alta, el comportamiento de TCP se ve perjudicado aumentando así el retardo de las comunicaciones [8]. Por lo tanto, TCP no es capaz de garantizar el bajo retardo que requieren algunas aplicaciones IIoT [9]. Para mejorar el comportamiento ofrecido por la capa de transporte, se han desarrollado varias alternativas, muchas de ellas pequeñas modificaciones o actualizaciones de TCP, como pueden ser STCP [10], Real-Time TCP [11] o Network Coded TCP [12], entre otras. QUIC aparece como una alternativa más disruptiva, siendo un protocolo de transporte que tiene como objetivos principales reducir la latencia en el establecimiento de la conexión, desarrollo de nuevas características y dotar de seguridad a las comunicaciones habituales en aplicaciones HTTP [13].

En este trabajo se propone la utilización de QUIC como protocolo de transporte para tráfico MQTT, analizando los beneficios temporales de esta combinación en entornos IoT reales, comparándola con las soluciones más empleadas actualmente. Por lo tanto, las principales contribuciones de este trabajo son las siguientes:

- Integración y optimización del socket de QUIC en las implementaciones del broker y cliente de MQTT en el lenguaje de programación GO.
- Código *open-source* en un repositorio público de *github*.
- Análisis del comportamiento de MQTT sobre QUIC, utilizando contenedores de Linux y el simulador de eventos ns-3 para emular distintas tecnologías inalámbricas.

La estructura del artículo es la siguiente: la Sección II describe los antecedentes y el trabajo relacionado. La Sección III ilustra el proceso de implementación de MQTT y QUIC en GO, mientras que la Sección IV explica la configuración de las redes y las conexiones que se utilizan para llevar acabo los experimentos. La Sección V muestra los escenarios y experimentos, comenta los resultados obtenidos y compara el esquema propuesto con la solución tradicional TCP/TLS/MQTT. Por último, la Sección VI recoge las conclusiones finales, así como líneas futuras de investigación.

II. ESTADO DEL ARTE

QUIC es un protocolo de transporte desarrollado originalmente por Google Inc. [13] y recientemente estandarizado por el IETF [14][1]. Además de abordar algunas de las limitaciones de TCP, QUIC proporciona algunos beneficios adicionales que pueden ser relevantes para los escenarios de IIoT, ya que es capaz de asegurar latencias más bajas y, a su vez, ofrecer un servicio fiable y seguro. Por su parte, MQTT es uno de los protocolos de aplicación más populares en el ámbito de las redes IIoT. Esta sección describe el funcionamiento básico de los protocolos MQTT y QUIC, para así entender la motivación de combinar ambos y el impacto que dicha combinación puede llegar a tener en el entorno IoT industrial.

A. MQTT

MQTT [4] es un protocolo basado en el modelo de publicación-suscripción. Gracias a su programación y al bajo ancho de banda de red requerido, ha sido ampliamente utilizado para conectar pequeños dispositivos en una variedad de industrias desde 1999. La versión 3.1.1 fue presentada en 2014 por IBM, y fue estandarizada por ISO y OASIS. Además, la versión 5.0 se ha estandarizado, aunque la versión 3.1.1 sigue siendo la más ampliamente usada [5].

En MQTT existen tres roles: *subscriber*, *publisher* y *broker*. Los *publisher* suelen ser pequeños sensores que generan (perciben/miden) información y la publican en un *broker* común con *topics* específicos. El *subscriber* consume los datos producidos por los *publisher*, de manera que recibe todos los mensajes que haya enviado cualquier *publisher* sobre el *topic* concreto al que está suscrito. Tanto los *publisher* como los *subscribers* pueden considerarse clientes en la topología de red correspondiente. El elemento clave de MQTT es el servidor *broker*, que gestiona las suscripciones. Todos los mensajes publicados en la red se envían al *broker*, que se encarga de distribuir la información a los *subscribers* correspondientes. El *broker* también tiene en cuenta los distintos niveles de calidad de servicio (QoS de sus siglas en inglés) para los clientes, así como las posibles retransmisiones. Aunque los clientes sólo interactúan con un *broker*, el sistema puede contener varios servidores que intercambian datos sobre los *topics* de sus *subscribers* actuales.

Una de las principales ventajas de MQTT es el aislamiento que consigue entre *publishers* y *subscribers*. Esto facilita la implementación de los distintos roles de cliente en dispositivos de baja capacidad computacional, pudiendo interactuar entre sí mediante MQTT. Otro elemento positivo es que el *publisher* puede enviar nuevo contenido siempre que esté disponible, desvinculando así la relación temporal entre el interés de un nodo y la publicación de la información.

B. QUIC

Los dos retos que pretende abordar QUIC, y que afectan principalmente al tráfico web, son la minimización de la latencia para una mejor experiencia de usuario, y proporcionar comunicaciones seguras [15], [16]. En la Figura 1 se muestra la pila de protocolos propuesta con QUIC, comparándola con la solución tradicional TCP/TLS transportando tráfico HTTP.

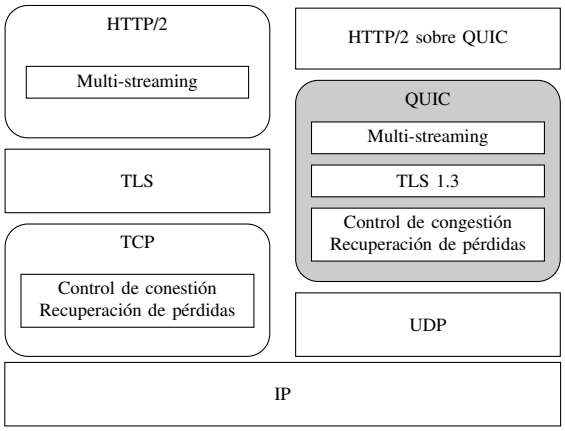


Fig. 1. Arquitectura tradicional de la capa de transporte con TCP/TLS frente a la propuesta con QUIC.

QUIC reduce la latencia en el *handshake*, estableciendo una conexión segura con el protocolo Transport Layer Security (TLS), versión 1.3, en solo un Round Trip Time (RTT), o incluso con cero RTT, si los extremos de la comunicación se han conectado previamente [16]. Por su parte, TCP implementa un intercambio de mensajes inicial

[1]Los RFC de QUIC han sido publicados después de finalizar los experimentos descritos en este trabajo. Por tanto, la versión de QUIC a la que se refiere el artículo es el draft 27, que es la que se empleó a lo largo de todo el trabajo.

denominado *3-way handshake* lo que hace que siempre aparezca un retardo inicial (en el primer segmento de datos) de un RTT [6]. Además, TCP necesita el protocolo TLS para encriptar los datos y poder garantizar así comunicaciones seguras. El establecimiento de la conexión de TLS 1.2 acarrea 2 RTT [17] y la versión 1.3 1 o 0 RTT [18]. Así, al utilizar TLS sobre TCP daría lugar a un *handshake* global de 3 o 2 RTT, respectivamente.

En una conexión QUIC la información está organizada en *streams*, evitando así el retardo causado por la pérdida de paquetes iniciales que bloqueen el resto. Cuando un paquete se pierde, solo los *streams* con datos en dicho paquete se bloquean esperando a la retransmisión del mismo mientras el resto continua [19]. También cabe destacar que QUIC reduce la latencia mediante sus mecanismos de detección de pérdida, incluyendo *early retransmits* y *tail loss probes* [19].

La principal diferencia con los mecanismos de detección de pérdidas utilizados por TCP es que QUIC no retransmite todos los paquetes con el mismo número de secuencia. QUIC, además de utilizar reconocimientos para confirmar la correcta recepción de los paquetes, usa un margen temporal para detectar pérdidas de cola (*probe timeout*). Para aplicar los mecanismos de recuperación que implementa QUIC se distinguen los siguientes supuestos de pérdida de paquete [19]:

- Si no ha llegado su ACK y fue enviado antes de otro paquete posterior ya reconocido.
- Si se mandó previamente con un margen temporal de $9/8 \cdot$RTT o su número de secuencia es 3 veces más pequeño que el del último paquete confirmado.
- Si se alcanza el *probe timeout* desde su envío y no hay ningún paquete posterior reconocido.

El tráfico TCP en Internet se supervisa y, en su caso, se mejora a través de *middleboxes*. Estos dispositivos analizan el contenido de los paquetes a nivel de transporte, y los modifican para conseguir un comportamiento óptimo de la red, por ejemplo adecuando la retransmisión de paquetes [20]. Es importante resaltar que cualquier mejora que se quiera hacer en TCP conlleva una actualización de los *middleboxes* correspondientes ya que, en caso contrario, estos descartarían los segmentos TCP que no puedan procesar. Esto dificulta la actualización de TCP, ya que se implementa en el *kernel* de estos equipos.

QUIC fue diseñado para suplir la rigidez que presenta TCP a la hora de introducir mejoras [13], [20]. Al estar implementado sobre UDP, los *middleboxes* consideran los paquetes QUIC como parte del *payload* de los datagramas UDP. La encriptación de la cabecera y *payload* de QUIC previene la interferencia de los *middleboxes* en el tráfico QUIC en el futuro. La decisión de diseñar e implementar QUIC a nivel de usuario le confiere una mayor flexibilidad, con mayor libertad en términos de capacidad computacional, más interacción con servidores y facilita considerablemente la actualización del protocolo [13]. Sin embargo, este último punto no es de obligado cumplimiento, por lo que QUIC también podría integrarse en el *kernel* para mejorar todavía más su rendimiento [21].

QUIC incluye un intercambio de mensajes donde se negocia la versión a utilizar, habilitando la coexistencia de diferentes modificaciones del mismo. Esto simplifica la actualización del protocolo, y habilita su ampliación y optimización. Por ejemplo, QUIC podría ser ampliado para incluir las demás extensiones como *plugins* permitiendo así que un *endpoint* de la red comparta la funcionalidad que otro no tenga [22].

C. QUIC en entornos IoT

QUIC se está empezando a incluir en las pilas de protocolos para soluciones IoT [23], [24] aunque todavía no existen muchos trabajos que aborden su evaluación en estos escenarios. Liri et al. valoraron QUIC como protocolo IoT en [25], revelando que QUIC, como sustituto de protocolos IoT más tradicionales, no proporciona el rendimiento del protocolo Constrained Application Protocol (CoAP) [26]. Sin embargo, el comportamiento de QUIC en entornos con cierta inestabilidad es comparable al de MQTT-Sensor Networks, variante de MQTT para dispositivos con baja capacidad. Los autores sugieren que una implementación de QUIC más simplificada y optimizada podría ser una alternativa frente a CoAP.

Kumar y Dezfouli estudiaron el comportamiento de la implementación de QUIC creada por Google en escenarios IoT [27]. Compararon el comportamiento de MQTT sobre QUIC y TCP en entornos de simulación acotados, usando dispositivos de baja capacidad computacional, Raspberry Pi 3B. El análisis se basó en el estudio de la sobrecarga de paquetes en el establecimiento de la conexión, el efecto en la latencia al introducir pérdidas aleatorias de paquetes, el uso de la memoria y procesador cuando uno de los *endpoints* rompe la conexión, y el rendimiento durante la migración de la conexión. Sus resultados muestran que QUIC supera a TCP en varios aspectos, e identifican algunas características que deberían mejorarse para un mejor rendimiento en escenarios IoT.

Lars Eggert ha analizado recientemente la viabilidad de integrar QUIC en equipos IoT [28]. En un primer estudio, usó dispositivos más optimizados que los que se utilizaron en [27]: Particle Argon y ESP32-DevKitC V4. Para disminuir el uso de memoria de estos dispositivos de baja capacidad, eliminó algunas de las funcionalidades de QUIC que podían considerarse poco prácticas para entornos IoT. Después de evaluar el uso de memoria y el consumo de energía de estos dispositivos, utilizando esta implementación optimizada de QUIC, se concluye [28] que es una buena opción para dispositivos *edge*.

En este trabajo se analiza el comportamiento de QUIC tal y como se define en [15] sobre escenarios IoT. En lugar de estudiar la capacidad de adaptación de QUIC en dispositivos *edge*, como en el caso de [28], se evalúa la reducción de latencia que se consigue al emplear QUIC. Se consideran diferentes escenarios IoT donde los equipos se envían mensajes MQTT sobre QUIC y TCP. Se mide el tiempo que tardan dichas comunicaciones y se compara el comportamiento de MQTT con ambos protocolos de transporte. Se contemplan canales libres de errores y otros

en los que se inyectan pérdidas con valores de Frame Error Rate (FER) realistas, sobre distintas tecnologías inalámbricas como WiFi, 4G/LTE y enlaces satelitales.

III. IMPLEMENTACIÓN

La implementación de QUIC que se ha utilizado en este trabajo se basa en la versión draft del IETF, y está desarrollada en el lenguaje de programación GO, *quic-go*[2]. El cliente y servidor de MQTT se basan en el código *open-source* de Eclipse Paho[3] y VolantMQ[4] respectivamente. Al igual que *quic-go*, estas aplicaciones también están desarrolladas en GO. Soportan la especificación al completo de MQTT, concretamente las versiones 3.1 y 3.1.1. Eclipse Paho implementa una librería que permite conectar el cliente MQTT con el broker VolantMQ usando TCP, TLS o *WebSocket*. En base a estas implementaciones del cliente y broker MQTT, se ha integrado QUIC en ambos proyectos. La Figura 2 muestra un resumen de los cambios realizados sobre las implementaciones originales (*net.go* y *quic_udp.go*), para habilitar el uso de MQTT sobre QUIC.

En el caso del cliente, todas las funcionalidades que se utilizan en la capa de transporte, abrir o cerrar la conexión o enviar y recibir paquetes, están gestionadas desde la interfaz *net.go*. Gracias a esto, la integración del socket de QUIC se agiliza, ya que solo requiere cambios mínimos en el código original para integrar conexiones basadas en TCP, TCP+TLS y WebSocket. En lo que se refiere a la integración de las conexiones QUIC para MQTT, la interfaz *net.go* llama a la interfaz *client.go*, implementada en *quic-go*. En dicha implementación, se soporta la funcionalidad de 0-RTT para el establecimiento de la conexión, a través de la función *DialAddrEarly* (*client.go*). Esta función, además de encargarse de abrir una sesión con el servidor, permite mandar datos antes de que el *handshake* de QUIC finalice reduciendo la latencia en el inicio de la conexión. Esta implementación de MQTT con QUIC es *open-source* y está disponible y accesible en un repositorio público[5].

La implementación *open-source* del *broker* con QUIC también está disponible en un repositorio git[6]. En este caso, las funciones de la capa de transporte están implementadas usando la interfaz *transport/conn.go*. La implementación original presenta algunas restricciones y, debido a la incompatibilidad entre las interfaces de conexión de TCP y QUIC, se ha decidido que esta implementación solo soporte el socket de QUIC. Por lo tanto, se integra la interfaz *quic_udp.go*, la cual llama al *listener* de la interfaz *server.go* de *quic-go*. La función que se utiliza para habilitar en el lado del servidor el 0-RTT es *ListenAddrEarly*. Esta función habilita que un cliente que se haya conectado al *broker* previamente pueda utilizar la información almacenada en la caché de esa sesión, y así en el restablecimiento de la conexión no tengan que negociar nuevamente todos los parámetros. De esta forma, el intercambio de datos sucede antes de que finalice el *handshake*.

La Figura 2 muestra un esquema de la comunicación entre cliente y *broker*. El *broker* estará escuchando en el socket que se le especifique. Se ejecutará la implementación original si se requiere TCP/TLS o la modificación realizada en el marco de este trabajo si se prefiere QUIC. En el caso de realizarse una conexión sobre QUIC, el servidor utilizará la función *ListenAddrEarly()* y el cliente establecerá la sesión a partir de *DialAddrEarly()*. De esta manera, ambos aceptarán utilizar el mecanismo de 0-RTT. Tras crearse la sesión, el cliente abrirá un *stream* a través de *OpenStreamSync()* y el *broker* la aceptará mediante la función *AcceptStream()*.

Por último, cabe destacar que se han identificado algunos aspectos que evidencian la falta de optimización de la implementación de QUIC. Los protocolos de transporte, TCP en particular, son capaces de combinar múltiples paquetes de protocolos de nivel superior en un único paquete de capa de transporte (*piggybacking*). Uno de los problemas que se han encontrado es la imposibilidad de analizar aquellos escenarios que conllevan una transmisión de datos a ráfagas de los *publishers*. Esto se debe a que la versión que se está utilizando en este trabajo de *quic-go* (v0.15.1) no es capaz de agrupar múltiples paquetes de MQTT aunque pertenezcan al mismo *stream* de datos.

IV. ENTORNO DE SIMULACIÓN

Como herramienta de evaluación, para estudiar el rendimiento de MQTT sobre QUIC, se ha empleado el simulador de redes basado en eventos discretos ns-3. En concreto, entre todas las funcionalidades de ns-3 se hace uso de la posibilidad de conectar aplicaciones en tiempo real sobre contenedores LXC de Linux, mediante una red simulada. Se ha generado un entorno de simulación basado en el escenario propuesto en la Figura 3, que muestra dos contenedores Linux conectados mediante ns-3. Esto permite emular varias tecnologías inalámbricas modelando dos parámetros principales: el ancho de banda y el retardo.

Un contenedor ejecuta la aplicación del cliente, compuesta por un *publisher* y un *subscriber*, mientras que el otro hace las veces de broker/servidor. ns-3 interpreta a los contenedores Linux como nodos fantasma conectados mediante una red CSMA a otro nodo, considerado como router. Para asegurar que el cuello de botella se encuentra en el enlace inalámbrico, ya que es el elemento de interés en el estudio, se configura una capacidad alta (10Gbps) en los enlaces entre los nodos y los contenedores. Los routers están conectados mediante un enlace punto a punto (P2P), que se ajusta en función del ancho de banda, retardo y la tasa de pérdidas.

La Tabla I muestra los parámetros empleados para modelar las diferentes redes propuestas. Los buffers han

[2]Implementación de QUIC https://github.com/lucas-clemente/quic-go versión v0.15.1.

[3]Cliente MQTT Eclipse Paho, https://github.com/eclipse/paho.mqtt.golang versión v1.2.0.

[4]Broker MQTT, https://github.com/VolantMQ/volantmq versión v0.4.0-rc6.

[5]Cliente Eclipse Paho MQTT con el socket de QUIC, https://github.com/pgOrtiz90/paho.mqtt.golang

[6]Servidor/*broker* MQTT con soporte de QUIC, https://github.com/fatimafp95/volantmq_2

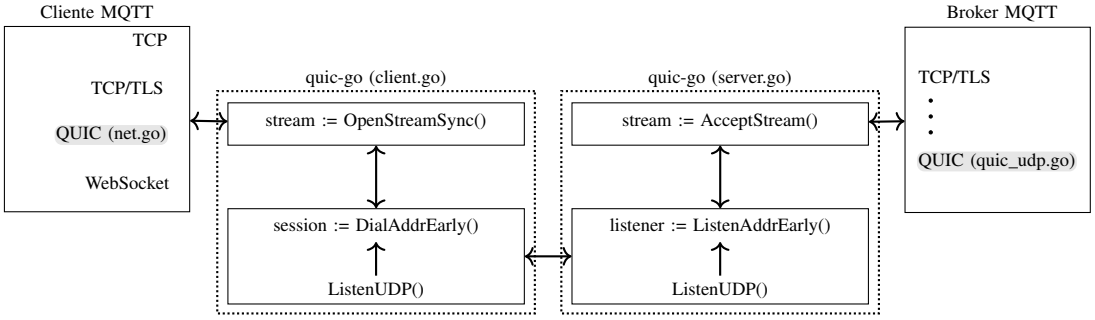


Fig. 2. Esquema de la integración de QUIC como protocolo de transporte para servicios basados en MQTT.

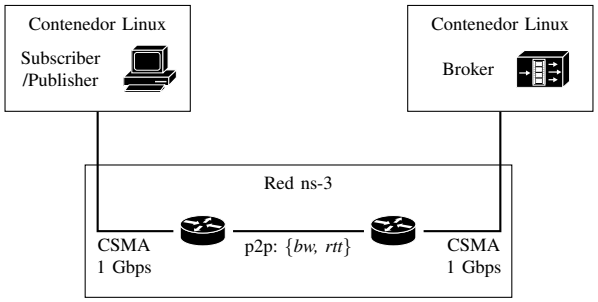


Fig. 3. Escenario inicial con dos contenedores Linux conectados por una red P2P: el contenedor de la izquierda ejecuta el cliente MQTT y el de la derecha el broker.

Tabla I
PARÁMETROS DE RED PARA EMULAR DISTINTAS TECNOLOGÍAS.

	Tipo1	**Tipo2**	**Tipo3**
	WiFi	**4G/LTE**	**Satélite**
Capacidad [Mbps]	20	10	1.5
RTT [ms]	25	100	600
Tasa de pérdidas [%]	[0, 1, 2, 3, 5, 10]		

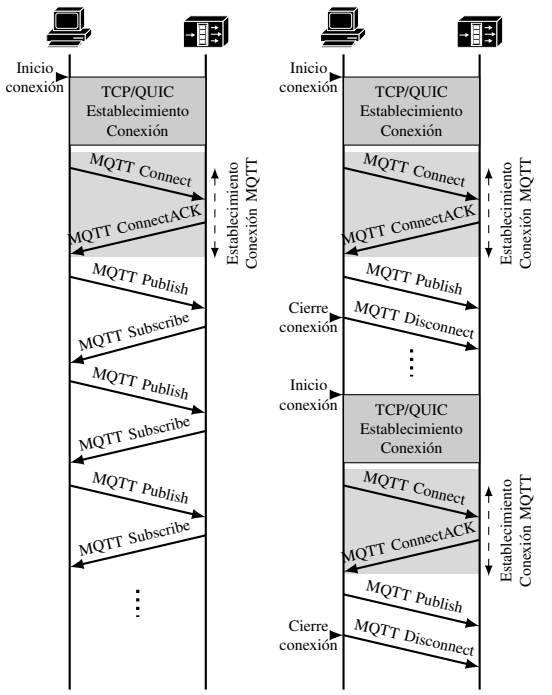
sido configurados a un Bandwidth Delay Product (BDP), teniendo aunque hemos tenido en consideración el denominado *bufferbloat effect* [29]. Este efecto se produce al tener tamaños de buffer superiores al BDP, algo característico en muchas redes móviles.

Por otro lado, ns-3 permite el uso de redes WiFi para conectar contenedores Linux, permitiendo analizar el rendimiento conjunto de MQTT y QUIC sobre un canal compartido. ns-3 proporciona herramientas para configurar la capa MAC y física, permitiendo fijar así el modelo de tasa de error. La interfaz que nos permite controlar este modelo de tasa de error en Wi-Fi calcula la probabilidad de recibir correctamente los paquetes sobre la capa física. Sin embargo, para emular las mismas condiciones del primer escenario, se ha modificado dicha interfaz, permitiendo controlar la tasa de pérdidas mostrada en la Tabla I, independientemente del modelo de interferencias empleado.

La configuración realizada está basada en una red inalámbrica ad-hoc. Dicha red conecta varios contenedores Linux como *publishers* MQTT al *broker*, el cual a su vez está conectado a un *subscriber* a través de una red similar.

V. RESULTADOS

En este trabajo se ha extendido la experimentación descrita en [30] para obtener un análisis más completo del comportamiento de MQTT sobre QUIC.



(a) Escenario con una única conexión MQTT.

(b) Escenario con múltiples conexiones MQTT.

Fig. 4. Escenarios implementados en la fase de experimentación.

En el primer escenario, que se muestra en la Figura 4a, se mandan 100 mensajes MQTT desde el *publisher* hasta el *broker*, el cual reenvía el mensaje correspondiente al subscriber. Se ejecuta cada experimento 50 veces para garantizar la validez estadística de los resultados. Se establece una sola conexión MQTT y no se cierra hasta que termine la transmisión de todos los mensajes. En este caso, el servicio que ofrece MQTT sigue un comportamiento *stop&wait*, donde el paquete i-*ésimo* solo se envía después de recibir la suscripción del anterior. Se analiza el tiempo necesario para transmitir todos los paquetes con TCP y QUIC, T_{TCP} y T_{QUIC} respectivamente. Para evaluar la reducción de T_{QUIC} frente a T_{TCP}, se define el parámetro ratio de finalización ξ como se muestra en la ec. 1

$$\xi = \frac{T_{\text{QUIC}}}{T_{\text{TCP}}} \tag{1}$$

Así, cuando $\xi < 1$, se puede decir que QUIC supera a TCP, ya que el tiempo que se alcanza para transmitir toda la información entre los endpoints es menor que el que se

obtiene con TCP. Este parámetro se mide sobre las tres redes que se configuran con los parámetros descritos en la Tabla I, con distintas tasas de pérdida de paquetes.

Para realizar un análisis más extenso, se profundiza en la metodología y configuración de [30] para generar escenarios más complejos. Considerando una arquitectura fog/cloud para IIoT y la configuración de la Figura 3, se añade otro nodo para separar los roles de los clientes de MQTT, donde el *publisher* y el *subscriber* se conectan al *broker* a través de distintas redes.

La Figura 5 muestra el ratio de finalización separando los clientes en distintos contenedores LXC. QUIC, en línea a lo analizado en [30], sigue presentando una menor latencia que TCP durante el intercambio de datos, especialmente sobre redes con RTT pequeños y tasa de pérdida alta gracias a sus mecanismos de recuperación de paquetes. Además, se deduce que cuando existen RTT elevados, la mejora de QUIC puede ser menos notable, como se puede ver en los enlaces satelitales. Sin embargo, QUIC sigue presentando una reducción de la latencia de aproximadamente 35% sobre redes Wi-Fi y un 5% de pérdidas.

Uno de los puntos fuertes del diseño de QUIC es la reducción temporal en el restablecimiento de la conexión. Para analizar esta mejora, y compararla con el esquema típico de TCP/TLS, se ejecuta el escenario 4b. Se cierra la conexión después de que el *publisher* mande un mensaje a un *topic* específico.

En [30] se muestra que aunque QUIC consiga mejorar las prestaciones de TCP, no aprovecha el 0-RTT. Después de analizar la implementación de *quic-go* se comprobó que había un problema con la gestión de los paquetes 0-RTT, que perjudicaba a QUIC. Esto se debía a que la implementación del servidor de QUIC no procesaba los paquetes 0-RTT antes de que el cliente retransmitiera datos 0-RTT en los paquetes 1-RTT. Después de detectar este inconveniente, los resultados mostrados en la Figura 6 avalan que QUIC maneja tiempos más cortos en el establecimiento de la conexión que el esquema tradicional TCP/TLS, concretamente cuando los nodos vuelven a comunicarse. Se ha ejecutado 50 veces la configuración de la Figura 3, separando los roles del cliente MQTT. Se han establecido dos conexiones para probar el impacto en la reconexión y se mide el tiempo desde que se establece la conexión inicial del *publisher* hasta que el *subscriber* recibe el segundo mensaje. En todos los casos QUIC supera en rendimiento a TCP/TLS, especialmente en enlaces con un RTT alto como los habituales en redes satelitales.

Teniendo en cuenta que el simulador ns-3 permite utilizar redes WiFi, así como los diferentes intercambios de tramas identificados en la Figura 4, se han utilizado varios clientes MQTT conectados a un *broker* mediante una red WiFi. Este escenario podría emular una red inalámbrica de sensores que se conectan a una arquitectura fog/cloud, en la que varios dispositivos edge generan datos que serán procesados por la capa fog. Finalmente, en el caso de ser necesario, serán enviados al cloud (rol del *subscriber*) a través de un enlace cableado con bajo RTT (25 ms) y un elevado ancho de banda.

La Figura 7 representa el parámetro ξ en el escenario descrito en 4a, al modificar la tasa de error y el número de retransmisiones que realiza la capa MAC. El experimento se ha realizado 50 veces para cada tasa de error y número de retransmisiones MAC. Con el objetivo de reducir el tiempo de simulación, se transmiten 100 paquetes MQTT desde el *publisher* al *subscriber*. La Figura 7 muestra que QUIC mejora el comportamiento de TCP, por lo que podría ayudar la latencia en redes WiFi y, en general, inalámbricas.

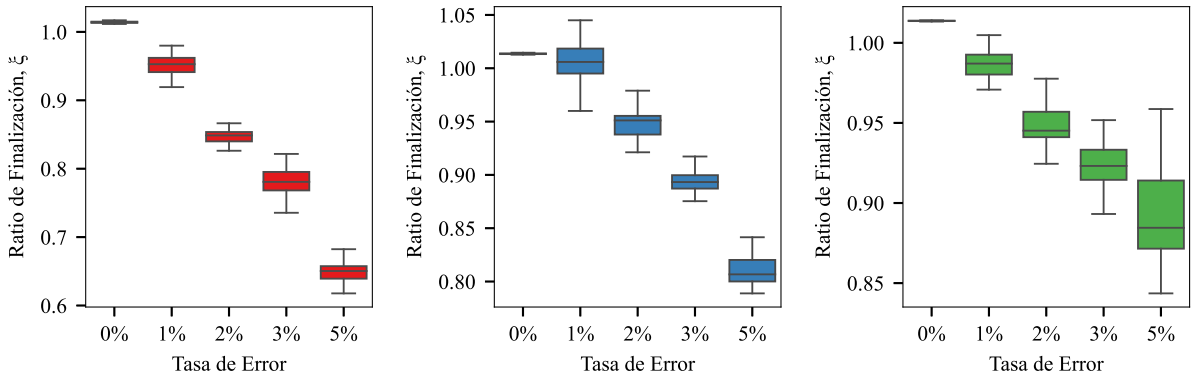
Por otro lado, se ha ejecutado un *subscriber* suscrito a un *topic* especifico variando el número de *publishers* en tres, cinco y ocho, donde uno de ellos publica información en dicho *topic*. La Figura 8 muestra el rendimiento en canales compartidos sobre el escenario descrito en la Figura 4a. QUIC mejora el rendimiento de TCP/TLS para todos los casos, mostrando un comportamiento más estable (con menor variabilidad) a medida que se incluyen más *publishers* generando datos.

VI. CONCLUSIONES Y LÍNEAS FUTURAS

En este paper se ha discutido y evaluado la alternativa de emplear QUIC como protocolo de transporte frente a la combinación TCP/TLS en aplicaciones industriales IIoT para reducir el retardo en estos entornos. En este tipo de escenarios la latencia es uno de los parámetros más críticos y, por tanto, este trabajo la analiza para comparar ambos protocolos. Las dos principales contribuciones de este paper son: en primer lugar la implementación basada en GO de MQTT sobre QUIC, y en segundo lugar, el análisis del rendimiento de dicha implementación en los escenarios IIoT emulados.

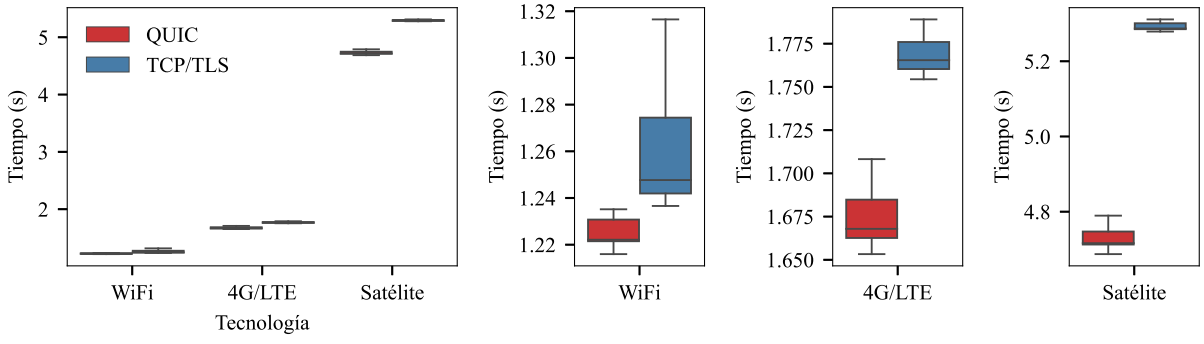
Por un lado, se han configurado tres tipos de redes en el simulador ns-3: Wi-Fi, 4G y satelital. Para ello se ha ajustado el ancho de banda y el retardo para emular diferentes tecnologías. Por otro lado, ns-3 permite también emular un escenario más realista, en el que clientes MQTT (things) publican mensajes a un *broker* (fog) haciendo uso de una red Wi-Fi. El *broker* a su vez transmite la información a un subscriber (cloud) conectado a través de una red P2P que emularía la conexión entre fog y cloud. Gracias a este simulador, y los contenedores Linux, se ha evaluado la combinación de QUIC y MQTT, comparando su comportamiento frente al ofrecido por la solución tradicional de MQTT/TLS/TCP.

A pesar de las restricciones derivadas de las implementaciones, se han diseñado dos escenarios complementarios que permiten analizar MQTT con QUIC y TCP. Por un lado, con el primer escenario se ha comprobado que QUIC excede en rendimiento a TCP en el intercambio de mensajes entre los nodos, especialmente sobre conexiones con un bajo RTT, volviéndose más evidente a medida que incrementa el número de paquetes que se pierden en la red. Se ha comprobado que el impacto del ancho de banda es casi insignificante, ya que se consideran paquetes de pequeño tamaño, típicos en IoT. Por otro lado, el segundo



(a) WiFi (b) 4G/LTE (c) Satélite

Fig. 5. Comportamiento de MQTT sobre QUIC y TCP para el escenario 4a.



(a) (b) (c) (d)

Fig. 6. 0-RTT con tres nodos: *publisher-broker-subscriber*. 6a muestra las tres configuraciones, y 6b, 6c y 6d el tiempo de transmisión para cada red.

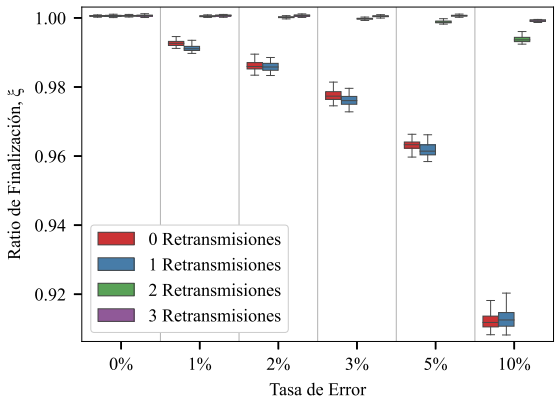


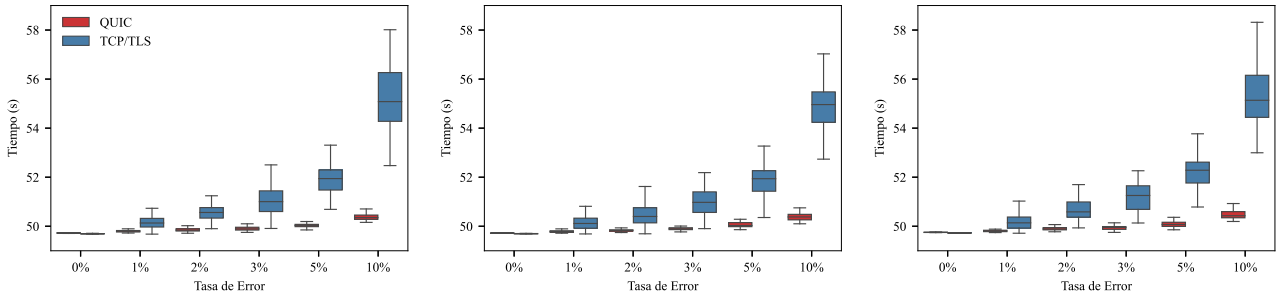
Fig. 7. *Boxplot* con el número de retransmisiones WiFi.

escenario fue concebido para conocer los beneficios del esquema 0-RTT que QUIC promueve. Los resultados muestran una clara reducción en la latencia durante el establecimiento de la conexión. Este logro podría ayudar a alcanzar comunicaciones más fiables en redes inalámbricas de entornos industriales. Por último, se ha evaluado QUIC en canales compartidos, configurando una red Wi-Fi en el simulador ns-3. Además, se ha concluido que QUIC presenta un comportamiento aceptable sobre entornos compartidos, como son las redes de sensores. Todos estos resultados se llevan a cabo para evaluar el rendimiento de QUIC frente a TCP en términos de latencia. QUIC ofrece tiempos más cortos, debido a los mecanismos de recuperación de pérdidas y a la funcionalidad 0-RTT.

Gracias al prematuro estado de desarrollo de QUIC se han encontrado aspectos interesantes que pueden ser aplicados, como añadir técnicas multipath en la implementación de *quic-go* o estudiar mecanismos de control para la congestión en IoT. Para garantizar la estabilidad y la disponibilidad de la red, es esencial evaluar los eventos de congestión e intentar evitarlos. Combinar multipath en QUIC junto con el algoritmo de control de congestión podría ser interesante, ya que los eventos de congestión se reducirían significativamente. La conexión multipath podría aliviar los cuellos de botella que se generan, redirigiendo el tráfico por rutas con menos carga. En este trabajo se considera que estas propuestas constituyen interesantes puntos de estudio para su uso en el entorno IIoT en los que la fiabilidad, la latencia y la capacidad son de vital importancia.

AGRADECIMIENTOS

Los autores agradecen la financiación del Programa de Doctorados Industriales de la Universidad de Cantabria (convocatoria 2020). El trabajo ha sido financiado por el Gobierno Vasco a través del programa Elkartek, y el proyecto DIGITAL (KK-2019/0009), y por la Agencia Estatal de Investigación, proyecto FIERCE: Future Internet Enabled Resilient smart CitiEs (RTI2018-093475-AI00).



(a) 3 *publishers* (b) 5 *publishers* (c) 8 *publishers*

Fig. 8. Comportamiento de MQTT sobre QUIC y TCP/TLS en canales compartidos con el escenario 4a.

REFERENCIAS

[1] M. Hermann, T. Pentek, and B. Otto, "Design Principles for Industrie 4.0 Scenarios," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 2016, pp. 3928–3937.

[2] G. Aceto, V. Persico, and A. Pescapé, "A Survey on Information and Communication Technologies for Industry 4.0: State-of-the-Art, Taxonomies, Perspectives, and Challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3467–3501, 2019.

[3] A. Varghese and D. Tandur, "Wireless requirements and challenges in Industry 4.0," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, nov 2014, pp. 634–638.

[4] A. Banks and R. Gupta, "MQTT version 3.1.1," International Organization for Standardization (ISO), Standard, 2014.

[5] A. Banks, E. Briggs, K. Borgendale, and R. Gupta, "MQTT version 5.0," Organization for the Advancement of Structured Information Standards (OASIS), Standard, 2019.

[6] J. Postel, "Transmission control protocol," Internet Requests for Comments, RFC Editor, STD 7, September 1981, http://www.rfc-editor.org/rfc/rfc793.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc793.txt

[7] W. Shang, Y. Yu, R. Droms, and L. Zhang, "Challenges in iot networking via tcp/ip architecture," *NDN Project*, 2016.

[8] Chonggang Wang, K. Sohraby, Bo Li, M. Daneshmand, and Yueming Hu, "A survey of transport protocols for wireless sensor networks," *IEEE Network*, vol. 20, no. 3, pp. 34–40, may 2006.

[9] J. Luo, J. Jin, and F. Shan, "Standardization of Low-Latency TCP with Explicit Congestion Notification: A Survey," *IEEE Internet Computing*, vol. 21, no. 1, pp. 48–55, 2017.

[10] R. Stewart, "Stream control transmission protocol," Internet Requests for Comments, RFC Editor, RFC 4960, September 2007, http://www.rfc-editor.org/rfc/rfc4960.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4960.txt

[11] C. Zhang and V. Tsaoussidis, "Tcp-real: Improving real-time capabilities of tcp over heterogeneous networks," in *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, ser. NOSSDAV '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 189–198.

[12] M. Kim, J. Cloud, A. ParandehGheibi, L. Urbina, K. Fouli, D. Leith, and M. Medard, "Network Coded TCP (CTCP)," *arXiv e-prints*, p. arXiv:1212.2291, 2012.

[13] A. Langley, J. Iyengar, J. Bailey, J. Dorfman, J. Roskind, J. Kulik, P. Westin, R. Tenneti, R. Shade, R. Hamilton, V. Vasiliev, A. Riddoch, W.-t. Chang, Z. Shi, A. Wilk, A. Vicente, C. Krasic, D. Zhang, F. Yang, F. Kouranov, and I. Swett, "The QUIC Transport Protocol," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication - SIGCOMM '17*. New York, New York, USA: ACM Press, 2017, pp. 183–196.

[14] J. Iyengar and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport," RFC 9000, IETF, Tech. Rep. 9000, May 2021, https://rfc-editor.org/rfc/rfc9000.txt. [Online]. Available: https://rfc-editor.org/rfc/rfc9000.txt

[15] ——, "Quic: A udp-based multiplexed and secure transport," Working Draft, IETF Secretariat, Internet-Draft draft-ietf-quic-transport-27, February 2020, http://www.ietf.org/internet-drafts/draft-ietf-quic-transport-27.txt. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-quic-transport-27.txt

[16] M. Thomson and S. Turner, "Using tls to secure quic," Working Draft, IETF Secretariat, Internet-Draft draft-ietf-quic-tls-27, February 2020, http://www.ietf.org/internet-drafts/draft-ietf-quic-tls-27.txt. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-quic-tls-27.txt

[17] T. Dierks and E. Rescorla, "The transport layer security (tls) protocol version 1.2," Internet Requests for Comments, RFC Editor, RFC 5246, August 2008, http://www.rfc-editor.org/rfc/rfc5246.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc5246.txt

[18] E. Rescorla, "The transport layer security (tls) protocol version 1.3," Internet Requests for Comments, RFC Editor, RFC 8446, August 2018.

[19] J. Iyengar and I. Swett, "Quic loss detection and congestion control," Working Draft, IETF Secretariat, Internet-Draft draft-ietf-quic-recovery-27, March 2020, http://www.ietf.org/internet-drafts/draft-ietf-quic-recovery-27.txt. [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-quic-recovery-27.txt

[20] M. Honda, Y. Nishida, C. Raiciu, A. Greenhalgh, M. Handley, and H. Tokuda, "Is it still possible to extend TCP?" *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*, pp. 181–194, 2011.

[21] P. Wang, C. Bianco, J. Riihijärvi, and M. Petrova, "Implementation and performance evaluation of the quic protocol in linux kernel," in *Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 227–234.

[22] Q. De Coninck, F. Michel, M. Piraux, F. Rochet, T. Given-Wilson, A. Legay, O. Pereira, and O. Bonaventure, "Pluginizing quic," in *Proceedings of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 59–74.

[23] A. Čolaković and M. Hadžialić, "Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues," *Computer Networks*, vol. 144, pp. 17–39, oct 2018.

[24] Nikshepa and V. Pai, "Survey on iot security issues and security protocols," *International Journal of Computer Applications*, vol. 180, no. 42, pp. 16–21, May 2018.

[25] E. Liri, P. K. Singh, A. B. Rabiah, K. Kar, K. Makhijani, and K. Ramakrishnan, "Robustness of IoT Application Protocols to Network Impairments," in *2018 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. IEEE, jun 2018, pp. 97–103.

[26] Z. Shelby, K. Hartke, and C. Bormann, "The constrained application protocol (coap)," Internet Requests for Comments, RFC Editor, RFC 7252, June 2014, http://www.rfc-editor.org/rfc/rfc7252.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc7252.txt

[27] P. Kumar and B. Dezfouli, "Implementation and analysis of QUIC for MQTT," *Computer Networks*, vol. 150, pp. 28–45, feb 2019.

[28] L. Eggert, "Towards securing the internet of things with quic," EasyChair Preprint no. 2434, EasyChair, 2020.

[29] H. Jiang, Y. Wang, K. Lee, and I. Rhee, "Tackling bufferbloat in 3G/4G networks," in *Proc. of the 2012 ACM conference on Internet Measurement Conference (IMC)*, 2012, pp. 329–342.

[30] F. Fernández, M. Zverev, P. Garrido, J. R. Juárez, J. Bilbao, and R. Agüero, "And quic meets iot: performance assessment of mqtt over quic," in *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2020, pp. 1–6.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Confiabilidad en la capa de transporte para la red de sensores antártica

Adrià Mallorquí, Agustín Zaballos, Alan Briones, Guiomar Corral
Grupo de Investigación en Internet Technologies and Storage (GRITS)
La Salle – Universitat Ramon Llull
Quatre Camins 30, Barcelona 08022
adria.mallorqui@salle.url.edu, agustin.zaballos@salle.url.edu, alan.briones@salle.url.edu, guiomar.corral@salle.url.edu

El proyecto de investigación SHETLAND-NET aspira a desarrollar un servicio de telemetría de *Internet of Things* (IoT) en la Antártida mediante la interconexión de *Wireless Sensor Networks* (WSN) a través de radioenlaces *Near Vertical Incidence Skywave* (NVIS) que conforman una *Long Fat Network* (LFN). Esta arquitectura presenta algunas propiedades típicas de las denominadas *challenging networks*, requiriendo una evaluación de la viabilidad de la solución propuesta y un análisis de qué protocolo de transporte puede aportar una mayor confiabilidad para este caso de uso. Para este propósito, se define y presenta un modelo de confiabilidad heterogéneo basado en capas. A través de extensivas simulaciones se valida el modelo, se comparan distintos protocolos de transporte y se evalúa la confiabilidad del sistema.

***Palabras Clave*- Protocolo de transporte, confiabilidad, Antártida, Long Fat Network, IoT, NVIS**

I. Introducción

Es bien sabido que en la Antártida se llevan a cabo múltiples estudios científicos de distintos ámbitos de investigación [1]. Durante las campañas de investigación, los investigadores se establecen temporalmente en bases antárticas, normalmente ubicadas en las zonas periféricas del continente antártico. Uno de los mayores retos de la Antártida es la falta de sistemas de telecomunicaciones convencionales [1], hecho que dificulta el despliegue de *Wireless Sensor Networks* (WSN), reduciendo las posibilidades de añadir mejoras a los estudios actuales (p. ej. mediante la comunicación entre dos bases remotas o automatización en la recolección de datos).

Para superar estas dificultades, el proyecto de investigación SHETLAND-NET continúa nuestro estudio del uso de radioenlaces de Alta Frecuencia (HF) usando la técnica *Near Vertical Incidence Skywave* (NVIS), proporcionando comunicaciones de bajo consumo en la Antártida. La señal transmitida rebota en la ionosfera, consiguiendo un *long backhaul link* con un área de cobertura de 250 km de radio y un ancho de banda de pocos kbps [2], [3]. Las redes con este tipo de enlaces se pueden clasificar como *Long Fat Networks* (LFN), caracterizadas por contener enlaces con un *Bandwidth Delay Product* (BDP) mayor a 12.500 bytes [4].

La tecnología NVIS se puede utilizar para interconectar bases científicas remotas, y nuestro objetivo es acabar desplegando una red *Internet of Things* (IoT) que interconecte WSN remotas [2]. Este despliegue en el campo lo llevaremos a cabo durante la siguiente campaña antártica. Sin embargo, las comunicaciones NVIS pueden ser propensas a errores según el estado de la ionosfera. Se pueden presentar situaciones típicas de las *challenging networks* [5], como por ejemplo conectividad intermitente, desconexiones puntuales extremo a extremo y tasas de error variables, que pueden degradar el rendimiento del servicio IoT en caso de emplearse una arquitectura de protocolo TCP/IP estándar.

Es necesario que, antes de la fase de despliegue, podamos estudiar e intentar anticipar la confiabilidad esperada del sistema de telemetría IoT que queremos desarrollar. De esta forma podríamos prever los posibles problemas de confiabilidad que podrían surgir y escoger qué contramedidas aplicar respectivamente.

Para este trabajo, nos centraremos en el caso de uso de la automatización de las estaciones *Ground Terrestrial Network-Permafrost* (GTN-P) [6]. Cada una de estas estaciones GTN-P se equipa con una placa Arduino que recoge 32 valores de distintos sensores cada hora. Estos valores deben ser enviados hasta el centro de control situado en una base científica. El Arduino de cada estación mandará estos valores hacia un concentrador Raspberry Pi 3B+ a través de comunicaciones *Long Range* (LoRa) en la red de acceso. Esta Raspberry actuará como *gateway*, reenviando los datos recibidos hacia el centro de control a través de los enlaces NVIS (red *backbone*) [2].

La fiabilidad de los enlaces NVIS depende mucho del estado de la ionosfera y de la actividad solar de forma que, durante la noche, no es posible mandar datos empleando la

misma frecuencia de transmisión que durante el día (prácticamente todos se perderían). Por esta razón, aplicamos una técnica de envío oportunista propio de las *Delay Tolerant Networks* (DTN) para enviar todos los datos recolectados durante la noche cuando el enlace NVIS pasa a estar disponible por la mañana. Cada concentrador debería haber recolectado 13 sets distintos de valores por cada estación GTN-P durante la noche. Se espera que durante este momento de envío oportunista la red se congestione debido a la cantidad de datos trasmitidos. Nuestro proyecto requiere que, de media, un mínimo de 9 de los 13 sets de datos de cada estación (aproximadamente un 70%) lleguen al centro de control correctamente. Creemos que es necesario evaluar qué protocolo de transporte se utilizará para este envío de una gran cantidad de datos [4], ya que puede influir en el rendimiento y la confiabilidad del servicio, especialmente en situaciones de congestión. Por este motivo, queremos estudiar y comparar el uso de distintos protocolos de transporte sobre la LFN modelando el escenario y el caso de uso en el simulador Riverbed Modeler.

El resto del artículo se estructura de la siguiente forma. En la sección II se describe el trabajo relacionado en protocolos de transporte y confiabilidad de los sistemas. En la sección III se presenta nuestra propuesta de modelo para medir y evaluar la confiabilidad de un sistema. En la sección IV se detallan los test y simulaciones ejecutados. En la sección V se discuten los resultados obtenidos. Finalmente, en la sección VI se comentan las conclusiones.

II. Trabajo relacionado

A. Protocolos de transporte

El rendimiento de los protocolos de transporte ha sido un tema de discusión y desarrollo desde que Internet fue concebido [7]. Los protocolos de transporte tradicionales como *Transmission Control Protocol* (TCP) sufren un bajo rendimiento en ciertos tipos de redes, entre las cuales se encuentran las LFNs. El concepto de LFN y sus efectos sobre el rendimiento de TCP fueron definidos en la *Request for Comments* (RFC) 1072, que fue posteriormente actualizada por la RFC 1323 y finalmente la RFC 7323. Algunas variantes de TCP y otros protocolos de transporte desarrollados durante los últimos años han mejorado el rendimiento de la transmisión para LFN [8]. Sin embargo, estos interpretan que las pérdidas de paquetes siempre son originadas por una congestión en la red, reduciendo así la tasa de transmisión o ventana de congestión. Esta asunción no es cierta para redes inalámbricas, donde los paquetes también se pueden perder debido a la propia naturaleza del medio (p. ej. el *fading*, la movilidad o las interferencias) [9]. Reducir la ventana de congestión en estas situaciones también degrada el rendimiento de la transmisión, consiguiendo un *throughput* menor. Por este motivo, existen protocolos de transporte que implementan mecanismos para discernir entre pérdidas originadas por congestión y pérdidas por canal, para así solo reducir la ventana de congestión en el primer caso y aumentar el rendimiento del envío [10].

TCP CUBIC (RFC 8312) [7] es el protocolo de transporte más utilizado actualmente, ya que es la variante de TCP utilizada por defecto en la mayoría de sistemas operativos. Además, otros protocolos modernos como TCP BBR, Copa, Indigo y Verus [10] son capaces de conseguir un buen rendimiento, tal y como se ha demostrado en los extensivos test realizados por la plataforma Pantheon [8] de la Universidad de Standford. En el presente artículo nos apoyaremos en nuestro trabajo previo: el *Adaptive and Aggresive Transport Protocol* (AATP) y su evolución, el *Enhanced AATP* (EAATP), el cual incorpora un mecanismo para diferenciar el motivo de una pérdida de paquetes y otro mecanismo de *fairness* para adaptar la tasa de envío según las circunstancias estimadas de la red [4], [10].

B. Confiabilidad en Cyber Physical Systems

Un *Cyber Physical System* (CPS) se define como un sistema con capacidades físicas y computacionales integradas. Algunos ejemplos de CPS son los sistemas de control industrial, las redes eléctricas inteligentes y las WSNs, así como la mayoría de dispositivos que abarcan el IoT [11]. En la literatura se define la confiabilidad de un CPS, en términos generales, como la propiedad de un sistema para comportarse como se espera ante situaciones adversas [11]. Estas adversidades pueden venir derivadas de distintos motivos, como por ejemplo nodos defectuosos, errores bizantinos, comportamientos maliciosos y errores de red, entre otros. Por esta razón, se pueden encontrar distintos enfoques para medir y proporcionar confiabilidad que se basan en elementos dispares. Proponemos clasificar estos enfoques en las siguientes cuatro categorías, que serán la base para definir nuestro modelo a continuación:

1) Confiabilidad de los datos: se define como la posibilidad de poder confirmar la exactitud de un dato proporcionado por una fuente [12]. Existen varios métodos que pretenden detectar nodos defectuosos o fallos de lectura de valores. Por ejemplo, en [13] se utiliza una arquitectura de *fog computing* para detectar, filtrar y corregir datos anormales. En [14] se utiliza un sistema de detección de intrusión de datos para notificar datos erróneos provenientes de ataques maliciosos.

2) Confiabilidad de la red: se define como la probabilidad de que un paquete llegue a su destino a tiempo y sin ser alterado a pesar de las adversidades (p. ej. un error de enlace, la saturación del canal o ataques maliciosos, entre otros) [15]. La mejora de la confiabilidad de la red es un tema que se ha enfocado desde diferentes perspectivas como los protocolos de transporte, los protocolos de direccionamiento y control topológico [16] o las arquitecturas DTN [5].

3) Confiabilidad social: esta tendencia ha ganado atención desde la irrupción del concepto *Social Internet of Things* (SIoT) [17], [18]. Para la confiabilidad en SIoT, se utiliza la capacidad social de los nodos u objetos a la hora de establecer relaciones autónomamente para poder definir modelos de confianza y reputación que tengan en cuenta distintos factores evaluados por los propios nodos. En [19] se define un modelo subjetivo que considera factores como la capacidad computacional de los nodos, el tipo de

relación entre ellos, el número total de transacciones, la credibilidad de un nodo o el *feedback* recibido de otros nodos, entre otros. En [20] se propone otro modelo que define unos parámetros de entrada como el beneficio esperado ante un éxito, las pérdidas esperadas ante un fallo, el coste esperado y la importancia del objetivo, entre otros. En [21] se define un sistema descentralizado de autogestión de la confianza basado en un sistema de *feedback* reputacional donde se preserva la privacidad de las partes participantes.

4) Consenso: representa el estado en el que todos los participantes de un mismo sistema distribuido acuerdan un mismo valor o respuesta [22]. Los protocolos de consenso se pueden dividir en dos grandes bloques: los consensos *proof-based* y los consensos bizantinos. Los primeros están más orientados a tecnologías *blockchain*, donde todos los participantes compiten por minar un bloque, y los más utilizados son *Proof-of-Work*, *Proof-of-Stake* y sus variantes [22]. El principal inconveniente de estos protocolos para IoT es que la mayoría de dispositivos tienen un hardware sencillo y una capacidad computacional reducida, dificultando así las tareas de minado [22]. El segundo bloque de protocolos de consenso es el más clásico, orientado a la detección de errores bizantinos. Estos suelen implementar mecanismos cooperativos de votación para llegar a un acuerdo en vez de hacerlo de forma competitiva, consiguiendo un menor consumo de recursos. El principal problema de estos mecanismos es la baja escalabilidad debido a la gran cantidad de mensajes que deben intercambiar los nodos participantes. Los protocolos más utilizados en este ámbito son *Practical Byzantine Fault Tolerance* (PBFT), RAFT, PaXoS y Ripple, aunque más variantes han surgido a lo largo de estos últimos años [22].

III. Modelo de confiabilidad

Después de nuestra revisión bibliográfica, todos los trabajos que se pueden encontrar en la literatura se centran en áreas específicas de la confiabilidad, pero ninguno de ellos incluye las cuatro categorías. Este hecho puede llevar a no interpretar correctamente las causas subyacentes del nivel de confiabilidad, de forma que unas contramedidas no idóneas, o incluso contraproducentes, se podrían aplicar si las interdependencias entre las distintas categorías de confiabilidad no se consideran. Por este motivo, encontramos la necesidad de diseñar un modelo propio que exprese y permita trabajar adecuadamente el nivel de confiabilidad de un sistema e incluya las cuatro categorías mencionadas anteriormente. Este modelo debe ayudarnos a anticipar e identificar los posibles puntos débiles de nuestro sistema de telemetría IoT.

Nuestro modelo propuesto para medir la confiabilidad y evaluar el rendimiento de un CPS (en nuestro caso, un grupo de WSN remotas interconectadas para proporcionar un servicio de telemetría IoT en la Antártida) se basa en cuatro capas. El modelo está caracterizado por 1) dos capas base (Capa de Confiabilidad del Dato y Capa de Confiabilidad de la Red), 2) dos capas de extensión (Capa de Confiabilidad Social y Capa de Consenso) que incluye funcionalidades opcionales, y 3) las interacciones entre ellas. Postulamos que cada capa se caracteriza por su definición (alcance), cómo se mide la confiabilidad de esa capa (métrica), y cómo se puede mejorar el valor de esta métrica (contramedidas).

A. Capa de Confiabilidad del Dato

Esta capa tiene como objetivo confirmar la exactitud de los datos obtenidos por la fuente. Proponemos medir la confiabilidad de esta capa con la métrica *Faulty Sensing Ratio* (FSR), definida en la Ec. 1 como la proporción entre el número de valores sensados erróneamente (FSV) y el número total de valores sensados (TSV) en mismo periodo de tiempo. Cuánto más bajo sea el FSR, mejor será la confiabilidad de los datos.

$$FSR = \frac{FSV}{TSV} \quad (1)$$

Se pueden aplicar métodos autocorrectivos que intenten mitigar el efecto de datos anormales (FSV) en el nodo origen [13], [14]. Otros ejemplos son los *hashes*, *checksums*, y bits de paridad, entre otros (ver Fig. 1).

B. Capa de Confiabilidad de la Red

Esta capa es la responsable de asegurar que los paquetes lleguen a su destino a tiempo y sin ser alterados a pesar de las adversidades (p. ej. error de enlace o saturación del canal). Medimos la confiabilidad de esta capa con el *Packet Delivery Ratio* (PDR), definido en la Ec. 2 como el cociente entre el número total de paquetes correctamente recibidos por todos los nodos (Pr) y el número total de paquetes enviados por todos los nodos (Ps), durante el mismo periodo de tiempo. Cuanto más alto sea el PDR, mejor será la confiabilidad de la red.

$$PDR = \frac{Pr}{Ps} \quad (2)$$

En esta capa se pueden utilizar técnicas de codificación de la transmisión [23] para incrementar la robustez de la señal transmitida. También se utilizan habitualmente protocolos de encaminamiento y *Quality of Service* (QoS) para encontrar el mejor camino hacia un destino mediante los cuales se cuantifica la calidad o el rendimiento de los enlaces de la red [16]. Los protocolos de transporte y mecanismos de control de congestión también pueden mejorar la confiabilidad de la red [10]. En el caso de las *challenging networks* se utilizan arquitecturas y protocolos *overlay* DTN, como por ejemplo el *Bundle Protocol* [5].

C. Capa de Confiabilidad Social

Esta capa es la responsable de beneficiarse de la capacidad de los nodos de establecer relaciones sociales autónomamente para mejorar la confianza entre ellos y escoger con más probabilidad los valores correctos. Medimos la confiabilidad de esta capa con la métrica *Successful Transaction Rate* (STR), calculada como la proporción entre el número de transacciones satisfactorias

(STR) y el número total de transacciones (TT) en un mismo período de tiempo, tal y como se muestra en la Ec. 3. Una transacción l se considera satisfactoria cuando un nodo j espera recibir alguna información o dato v del nodo i antes de un tiempo máximo de recepción (Trx_{max}) y el valor recibido entra dentro de un intervalo esperado, de forma que el nodo j proporciona un *feedback* positivo sobre el nodo i ($f_{ij}^{l} = 1$). Cuanto mayor sea la STR, mejor será la confiabilidad social.

$$STR = \frac{ST}{TT} \tag{3}$$

Las contramedidas en esta capa suelen utilizar mecanismos de reputación para determinar en qué nodos confiar como proveedores y/o receptores de datos. Esta reputación se puede basar en experiencias previas, tanto por parte del nodo evaluador como de los nodos vecinos, que ayudan a construir una opinión sobre la confianza hacia otro nodo concreto [19], [21].

D. Capa de Consenso

Esta capa persigue alcanzar un estado donde todos los participantes de un grupo acuerden una misma respuesta o resultado conjunto (*General Agreement* o GA). Medimos la confiabilidad de esta capa con la métrica *Byzantine Node Tolerance* (BNT), definida como el cociente entre el número de nodos bizantinos (Nb) que se toleran en un grupo sin afectar al acuerdo llegado y el número total de nodos participantes en ese grupo (Nt), tal y como se muestra en la Ec. 4. Un nodo se considera bizantino si experimenta un fallo que le incapacite para comportarse como se espera o si no sigue el comportamiento esperado a propósito (malicioso). Cuanto mayor es la BNT, más probable es conseguir un GA correcto.

$$BNT = \frac{Nb}{Nt} \tag{4}$$

Se pueden utilizar multitud de mecanismos para obtener un GA descentralizado. Teóricamente, si el número de nodos bizantinos supera la mitad de los nodos totales participantes, cualquier mecanismo de consenso fallará al intentar conseguir un GA correcto [22]. El principal inconveniente de estos mecanismos es que los nodos participantes necesitan intercambiar un gran número de mensajes para conseguir el consenso, hecho que puede degradar el rendimiento de redes con poco ancho de banda.

E. Relaciones entre las capas de consenso

La Fig. 1 sintetiza los actores de nuestro modelo de confiabilidad. Los elementos azules forman parte de las capas base de nuestro modelo, mientras que los elementos naranjas forman parte de las capas de extensión. El objetivo fundamental es incrementar el STR para conseguir una mayor confiabilidad, de forma que esta es la métrica principal que utilizaremos en nuestras simulaciones. Hay tres factores que directamente ayudan a aumentar el STR: 1) Mitigar/tolerar errores bizantinos; 2) reducir el FSR; y 3) aumentar el PDR. Estos factores los consideramos subobjetivos de nuestro modelo. Cada uno de estos subobjetivos se pueden conseguir mediante la implementación de sus respectivas contramedidas, que

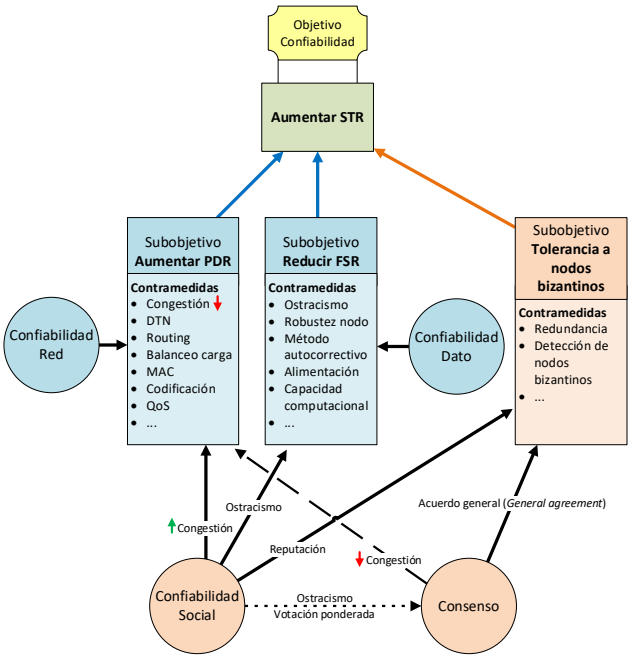


Fig. 1. Diagrama de relaciones del modelo de confiabilidad.

normalmente solo afectan a uno de los subobjetivos. Sin embargo, también detectamos dos acciones transversales que pueden afectar a más de un subobjetivo. Estas acciones transversales consisten en implementar las capas de extensión de nuestro modelo: la Capa de Confiabilidad Social y la Capa de Consenso.

En la Fig. 1, las flechas con línea continua indican una afectación positiva, las flechas con línea discontinua indican una afectación negativa, y las flechas con línea punteada indican una afectación incierta. Por ejemplo: el uso de la confiabilidad social puede reducir la congestión de la red, gracias al ostracismo de los nodos con peor reputación, que pueden dejar de generar tráfico, e intercambiar solo los datos de los nodos con más reputación. Este hecho también ayuda a mitigar los errores bizantinos, ya que se priorizará el uso los nodos de más confianza (que tendrán menos probabilidades de experimentar un error bizantino). Pero, por otro lado, la implementación de un mecanismo de consenso ayuda a tolerar los errores bizantinos gracias al GA al que llegan los nodos participantes de un mismo grupo. Sin embargo, la Capa de Consenso puede afectar negativamente al PDR, ya que introduce una cantidad considerable de tráfico extra que puede llegar a congestionar la red.

IV. Test y Simulaciones

Con el objetivo de 1) anticipar qué problemas pueden ocurrir durante la campaña antártica, 2) decidir qué protocolo de transporte utilizar para nuestro servicio, y 3) tener unas expectativas sobre los resultados del despliegue más precisas, aplicamos nuestro modelo de confiabilidad para medir y evaluar el caso de uso propuesto. Con esta finalidad, el escenario del caso de uso se ha modelado y testeado en el simulador Riverbed Modeler. A continuación, se detalla cómo se han modelado cada uno de los actores de nuestro caso de uso.

Tabla I
MODELO DE RED PARA LAS SIMULACIONES

Parámetro	NVIS	LoRa
Banda de transmisión	4.3 MHz	868 MHz
Ancho de banda	2.3 kHz	125 kHz
Bitrate	20 kbps	5,47 kbps
Rango de cobertura	Hasta 250 km	Hasta 30 km
Disponibilidad diurna (6am-5pm)	70%-100%	100% (LoS), 2%-100% (No LoS)
Disponibilidad nocturna (5pm-6am)	0%	100% (LoS), 2%-100% (No LoS)
Tamaño máximo de *payload*	242 bytes	140 bytes

Tabla II
PARÁMETROS DE LAS SIMULACIONES

Parámetro	Valor
Número de rondas por test	30
Duración	120 horas (5 días)
Pb_0	[1×10^{-3}, 2×10^{-3}, 4×10^{-3}, 8×10^{-3}, 1×10^{-2}, 2×10^{-2}, 4×10^{-2}, 8×10^{-2}, 1×10^{-1}]
k	5.7×10^{-5}
Protocolo de transporte	[BBR, Copa, CUBIC, EAATP, Indigo, Verus]
Modo de redundancia	[Ninguno, Social, Consenso]
Número de *gateways* NVIS	5
Número de *clusters* GTN-P por *gateway*	[8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096]
Número de estaciones GTN-P redundantes por *cluster*	[1-10]

En primer lugar, para el modelo de la red se han hecho dos modelos independientes para la red *backbone* (NVIS) y la red de acceso (LoRa), ya que usan tecnologías con características distintas. Estas redes se han caracterizado por separado tal y como muestra la Tabla I, partiendo de los resultados de [3] y [24] respectivamente. En segundo lugar, se han modelado los siguientes protocolos de transporte tal y como hicimos en nuestro trabajo previo [10]: BBR, Copa, CUBIC, EAATP, Indigo y Verus. Su rendimiento esperado se ha caracterizado según los resultados obtenidos en nuestro trabajo previo [10] y los test de la plataforma Pantheon [8].

En tercer lugar, es necesario modelar el comportamiento bizantino de los nodos. Según [25], la probabilidad que un nodo experimente un error bizantino, Pb_0, no es constante en el tiempo. De hecho, el incremento de esta probabilidad se puede asociar al desgaste del nodo, que también está relacionado con la descarga de la batería que lo alimenta. Siguiendo el modelo de [25], podemos asumir que el impacto del desgaste es lineal, como se define en la Ec. 5:

$$Pb(t) = Pb_0 + k \times t, \qquad (5)$$

donde Pb_0 es la probabilidad de que un nodo experimente un error bizantino en el instante $t = 0$ y k es el factor de desgaste. La probabilidad de que un nodo experimente un error bizantino incrementa a lo largo del tiempo hasta que la batería se vacía completamente en $t = t_d$. En este momento el nodo deja de responder y $Pb(t_d) = 1$. En las simulaciones se utilizan varios valores de Pb_0 para modelar el uso de distintos métodos correctivos.

A continuación, para el modelo de la confiabilidad social se ha utilizado una versión simplificada del modelo de confiabilidad objetiva de [19]. La simplificación para nuestro caso de uso sirve al escenario que se desplegará en la base antártica en el que todas las transacciones tienen la misma importancia, todos los nodos tienen la misma capacidad computacional, y el tipo de relación social entre todos los nodos es equivalente.

Finalmente, el protocolo de consenso se puede modelar sabiendo el tráfico extra que este añade a la red y el número de nodos bizantinos que este tolera (Nb). En nuestro caso, se ha escogido el protocolo PBFT [26] para implementar el consenso. El tráfico que este protocolo añade crece exponencialmente a medida que el número de nodos que participan en el mismo grupo de consenso (Nt) aumenta. Además, el número de nodos tolerados (Nb) se puede calcular según la Ec. 6:

$$Nb = \left\lfloor \frac{Nt-1}{3} \right\rfloor \qquad (6)$$

La Tabla II sintetiza los parámetros de nuestras simulaciones.

V. DISCUSIÓN DE LOS RESULTADOS

Tras ejecutar todas las simulaciones, se ha calculado la media del valor de STR obtenido por cada grupo de 30 test. Los resultados obtenidos tienen una desviación máxima del 0,68% con un intervalo de confianza del 99%. En nuestro caso de uso se pueden diferenciar tres modos de funcionamiento principales: el modo estándar, el modo de redundancia con la aplicación de la Capa de Confiabilidad Social, y el modo de redundancia con la aplicación de la Capa de Consenso. Por cada modo, se construye una matriz de N×M dimensiones con todas las combinaciones posibles de los parámetros de la simulación, donde N es el número de posibles combinaciones de *clusters* por *gateway* y estaciones GTN-P por *cluster* (100 en nuestro caso), y M es el número de distintos valores de Pb_0 posibles (9 en nuestro caso). Por cada punto de esta matriz y por cada protocolo de transporte se calcula el valor medio de STR obtenido. Si se unen todos los valores de STR calculados, podemos obtener una malla por cada protocolo de transporte. A esta malla la llamamos Malla de Confiabilidad. Las Fig. 2, 3 y 4 muestran la Malla de Confiabilidad para el modo estándar, el modo de redundancia con confiabilidad social y el modo de redundancia con consenso, respectivamente.

El eje "Probabilidad de error bizantino" tiene 9 puntos discretos, correspondientes a los valores Pb_0 de la Tabla II. El eje "Sensores redundantes × Número de *clusters*" tiene 100 puntos discretos, que son [1×2^N, 2×2^N, …, 10×2^N], donde N = [3, 4, …, 12]. Estos valores corresponden a los que se muestran en la Tabla II, filas 9 y 10. Aunque en las etiquetas del eje solo se muestran los valores iniciales de cada intervalo [1×8, 2×8, …, 10×8], dentro de cada intervalo se incrementa el número de *clusters* por *gateway* (p.ej. [1×8, 1×16, …, 1×4096].

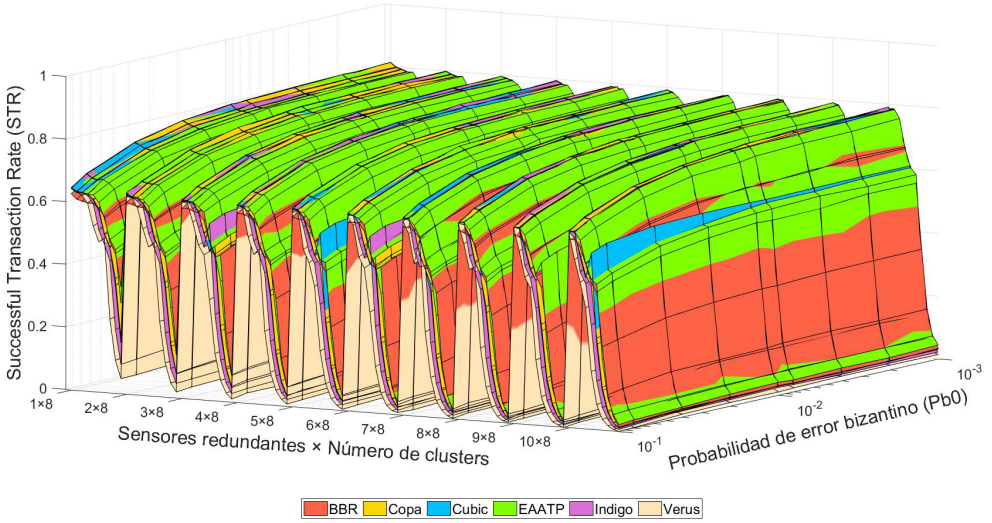


Fig. 2. Malla de confiabilidad (Estándar).

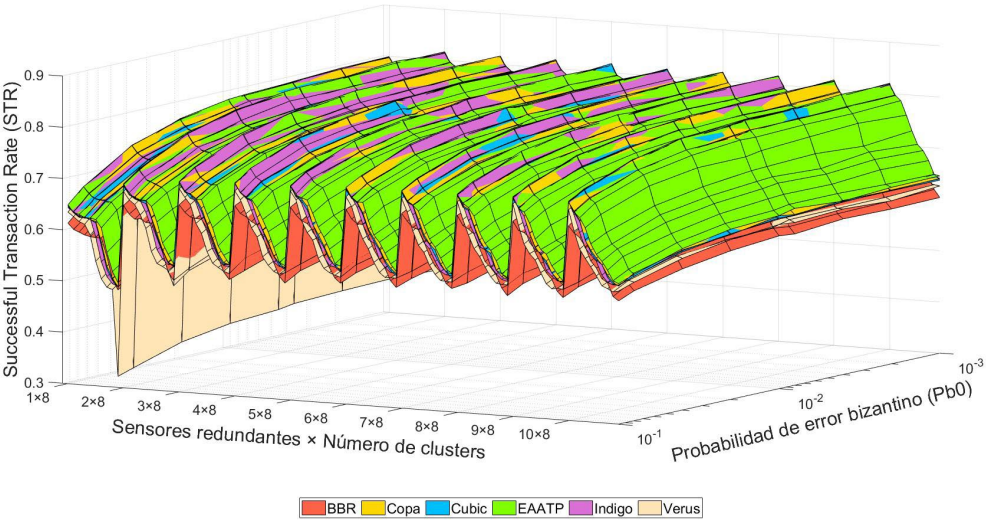


Fig. 3. Malla de confiabilidad (Social).

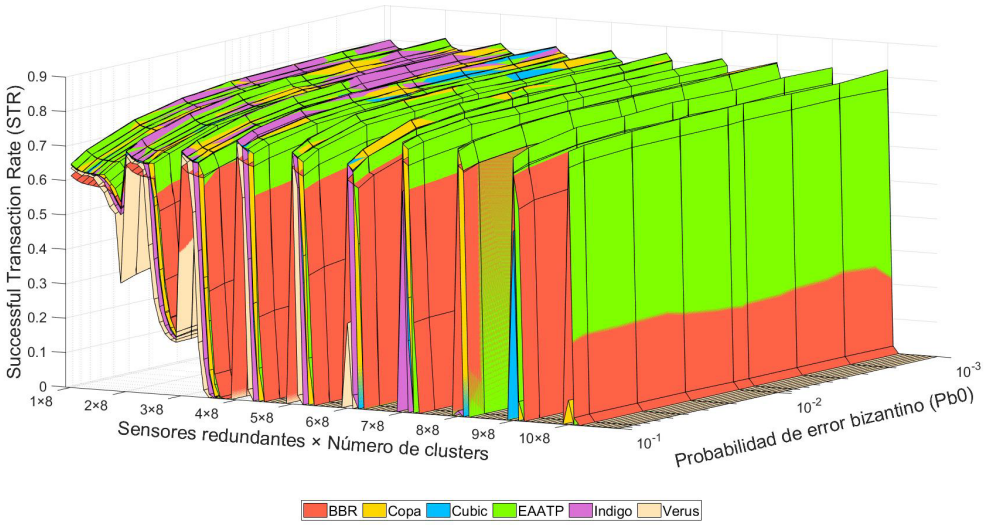


Fig. 4. Malla de confiabilidad (Consenso).

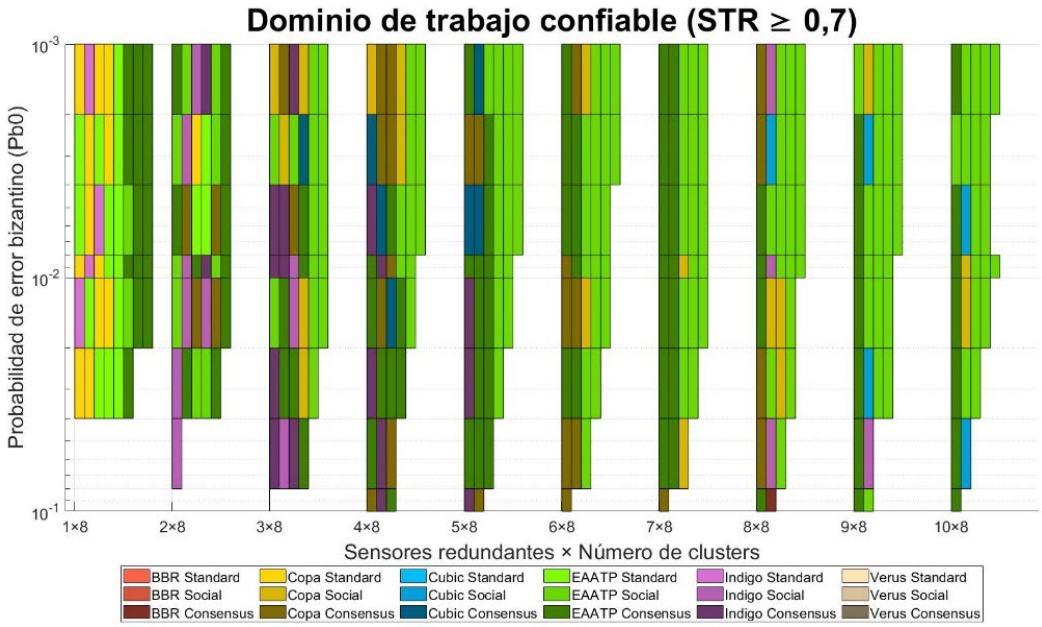


Fig. 5. Dominio de trabajo confiable para un *STR* mínimo de 0,7. Se muestra la mejor opción para cada caso.

En las Fig. 2, 3 y 4 podemos ver, por un lado, que los niveles de confiabilidad obtenidos son similares por todos los protocolos de transporte estudiados cuando hay pocos nodos en la red y, por lo tanto, esta se encuentra poco cargada. Sin embargo, 1) los niveles de *STR* obtenidos por BBR y Verus son ligeramente inferiores a los de sus competidores, y 2) Copa, Indigo y EAATP consiguen los valores de *STR* más altos para una baja carga de la red, aunque el dominio del EAATP como el protocolo con más confiabilidad crece a medida que la sobrecarga de la red aumenta, cuando se generan más pérdidas por congestión.

Por otro lado, también observamos que el modo de redundancia con confiabilidad social (Fig. 3) es el más robusto, ya que los valores de *STR* decrecen de una forma menos acentuada a media que la congestión y el número de nodos aumenta si lo comparamos con los otros casos (Fig. 2 y 4), manteniendo siempre unos valores superiores a 0,5. Además, se puede confirmar que en general tenemos más confiabilidad (mayores valores de *STR*) a medida que la Pb_0 decrece.

Nuestro modelo también se puede utilizar para visualizar el posible dominio de trabajo utilizable para implementar el servicio requerido, dado un valor mínimo de confiabilidad exigible. Nuestro caso de uso requiere un *STR* mínimo de 0,7, de forma que un promedio de 9 de los 13 valores leídos por cada sensor lleguen correctamente al centro de control, y así cumplir con el objetivo de [6]. La Fig. 5 muestra el dominio de trabajo para nuestro servicio de telemetría y un *STR* mínimo de 0,7. Por cada punto de la matriz, si no hay ninguna solución que consiga un *STR* igual o superior al mínimo deseado, este se deja en blanco, significando que la red no puede satisfacer los requerimientos del servicio bajo esas condiciones. Contrariamente, si una o más soluciones consiguen un *STR* igual o superior al mínimo deseado, ese punto se rellena con el color de la solución con un *STR* mayor. Desde esta perspectiva se puede apreciar un claro dominio del EAATP como la solución más confiable para la mayoría de los casos. Concretamente, el EAATP puede llegar a mejorar hasta un 7% sus competidores (p.ej. en el caso "7×8" con Pb_0=10^{-1}, EAATP logra una *STR* de 0,78 mientras que Verus obtiene 0,71), mientras que como máximo está un 0,5% por debajo cuando es superado por otro protocolo (p.ej., en el mismo caso que en el anterior, Copa logra una *STR* de 0,785).

Concluimos que el dominio del EAATP se debe a sus mecanismos de *fairness* y diferenciación de pérdidas. Por un lado, el mecanismo de *fairness* es capaz de repartir el ancho de banda del enlace entre varios flujos EAATP sin que se generen pérdidas por congestión. Por otro lado, el mecanismo de diferenciación de pérdidas es capaz de detectar si un paquete se ha perdido debido a una congestión de la red o a un error de canal, de forma que no reduce la tasa de envío en el segundo caso y consigue un mayor rendimiento. Estos factores le dan una ventaja competitiva al EAATP, ya que en este caso de uso se utiliza una mecánica propia de las DTN para almacenar todos los valores en un nodo intermedio durante la noche y mandarlos en bloque cuando el canal se encuentra disponible, hecho que congestiona la red. Además, el protocolo EAATP está concebido para utilizar la máxima capacidad posible de un enlace, hecho que puede ser muy relevante en nuestro caso de uso ya que las redes disponibles tienen un ancho de banda reducido y es crucial poder utilizarlo al máximo.

VI. Conclusiones

Este artículo continúa el trabajo del proyecto de investigación SHETLAND, el cual persigue el objetivo de diseñar e implementar un conjunto de WSN remotas en la Antártida que se interconectan mediante el uso de enlaces NVIS. Nuestro trabajo se ha centrado en analizar y comparar la confiabilidad de diferentes protocolos de transporte para nuestro caso de uso, el cual ofrece un servicio de telemetría IoT. Debido a las características de la ionosfera, los enlaces NVIS no funcionan de la misma forma durante la noche, motivo por el cual los valores adquiridos durante este periodo se almacenan

temporalmente en un nodo concentrador y se mandan en bloque cuando el canal se encuentra disponible, pudiendo congestionar la red. Para poder estudiar la viabilidad de esta arquitectura antes de implementar el servicio en la campaña antártica, y con el objetivo de comparar el rendimiento de varios protocolos de transporte, se ha definido un modelo para medir y evaluar la confiabilidad del sistema propuesto. Este modelo se compone de cuatro capas que pueden afectar a la principal métrica de confiabilidad, la STR, que mide el número de transacciones satisfactorias que llegan correctamente al centro del control y cuyo valor puede ser mejorado mediante la aplicación de contramedidas.

Se han analizado tres modos de funcionamiento y seis protocolos de transporte bajo distintas circunstancias con el simulador Riverbed Modeler. Los resultados muestran el dominio del protocolo EAATP como el más confiable para la mayoría de los casos (llegando a mejorar a sus competidores hasta un 7%), mientras que BBR y Verus son los menos confiables. Añadir redundancia de sensores y aplicar un método de confiabilidad social mejora la robusteza del servicio, consiguiendo valores de STR más altos y que en ningún caso son inferiores a 0,5 incluso en situaciones de carga elevada. Contrariamente, aplicar un mecanismo de consenso mejora la confiabilidad del sistema en situaciones de pocos nodos, pero parece contraindicado cuando estos aumentan dada la sobrecarga de tráfico que introduce. En un futuro, se pretende evaluar el uso de distintas técnicas DTN para mejorar la confiabilidad del sistema ante situaciones donde la caída de los enlaces NVIS no sea predecible.

AGRADECIMIENTOS

Este trabajo ha sido subvencionado por la "Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya", la Unión Europea y el Fondo Social Europeo [2021 FI_B1 00175], así como por la "Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) de la Generalitat de Catalunya" ("2017 SGR 977"). También se han recibido fondos del Ministerio de Ciencia, Innovación y Universidades del Gobierno de España, la Agencia Estatal de Investigación y el Fondo Europeo de Desarrollo Regional [RTI2018-097066-B-I00 (MCIU/AEI/FEDER, UE)]. Los autores quieren agradecer a La Salle Universitat Ramon Llull por el apoyo recibido durante la realización del proyecto.

REFERENCIAS

[1] M. C. Kennicutt *et al.*, "Delivering 21st century Antarctic and Southern Ocean science," *Antarct. Sci.*, vol. 28, no. 6, pp. 407–423, 2016, doi: 10.1017/S0954102016000481.

[2] J. Porte, J. M. Maso, J. L. Pijoan, and D. Badia, "Sensing System for Remote Areas in Antarctica," *Radio Sci.*, vol. 55, no. 3, pp. 1–12, 2020, doi: 10.1029/2019RS006920.

[3] J. Male, J. Porte, T. Gonzalez, J. M. Maso, J. L. Pijoan, and D. Badia, "Analysis of the Ordinary and Extraordinary Ionospheric Modes for NVIS Digital Communications Channels," *Sensors*, vol. 21, no. 6, p. 2210, 2021, doi: 10.3390/s21062210.

[4] A. Briones, A. Mallorquí, A. Zaballos, and R. M. de Pozuelo, "Adaptive and aggressive transport protocol to provide QoS in cloud data exchange over Long Fat Networks," *Futur. Gener. Comput. Syst.*, vol. 115, pp. 34–44, 2021, doi: 10.1016/j.future.2020.08.043.

[5] S. Bounsiar, F. Z. Benhamida, A. Henni, D. L. de Ipiña, and D. C. Mansilla, "How to Enable Delay Tolerant Network Solutions for Internet of Things: From Taxonomy to Open Challenges," *Proceedings*, vol. 31, no. 1, p. 24, 2019, doi: 10.3390/proceedings2019031024.

[6] M. Á. de Pablo Hernández *et al.*, "Frozen ground and snow cover monitoring in livingston and deception islands, antarctica: Preliminary results of the 2015-2019 PERMASNOW project," *Geogr. Res. Lett.*, vol. 46, no. 1, pp. 187–222, 2020, doi: 10.18172/cig.4381.

[7] S. Ha, I. Rhee, and L. Xu, "Cubic: a new TCP-friendly high-speed TCP variant," *ACM SIGOPS Oper. Syst. Rev.*, vol. 42, no. 5, pp. 64–74, 2008, doi: 10.1145/1400097.1400105.

[8] F. Y. Yan *et al.*, "Pantheon: The training ground for internet congestion-control research," in *Proceedings of the 2018 USENIX Annual Technical Conference, USENIX ATC 2018*, 2020, pp. 731–743.

[9] J. M. Chen, C. H. Chu, E. H. K. Wu, M. F. Tsai, and J. R. Wang, "Improving SCTP performance by jitter-based congestion control over wired-wireless networks," *Eurasip J. Wirel. Commun. Netw.*, vol. 2011, pp. 1–13, 2011, doi: 10.1155/2011/103027.

[10] A. Briones, A. Mallorquí, A. Zaballos, and R. M. de Pozuelo, "Wireless loss detection over fairly shared heterogeneous long fat networks," *Electronics*, vol. 10, no. 9, 2021, doi: 10.3390/electronics10090987.

[11] M. Crawford and E. Liongosary, *IIC Journal of Innovation*, vol. 9. The Industrial Internet of Things Consortium, 2018.

[12] N. Haron, J. Jaafar, I. A. Aziz, M. H. Hassan, and M. I. Shapiai, "Data trustworthiness in Internet of Things: A taxonomy and future directions," in *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, 2017, pp. 25–30, doi: 10.1109/ICBDAA.2017.8284102.

[13] G. Zhang and R. Li, "Fog computing architecture-based data acquisition for WSN applications," *China Commun.*, vol. 14, no. 11, pp. 69–81, 2017, doi: 10.1109/CC.2017.8233652.

[14] R. Fantacci, F. Nizzi, T. Pecorella, L. Pierucci, and M. Roveri, "False Data Detection for Fog and Internet of Things Networks," *Sensors*, vol. 19, no. 19, p. 4235, 2019, doi: 10.3390/s19194235.

[15] M. M. Hassan, A. Gumaei, S. Huda, and A. Almogren, "Increasing the Trustworthiness in the Industrial IoT Networks through a Reliable Cyberattack Detection Model," *IEEE Trans. Ind. Informatics*, vol. 16, no. 9, pp. 6154–6162, 2020, doi: 10.1109/TII.2020.2970074.

[16] H. P. Alahari and S. B. Yalavarthi, "A Survey on Network Routing Protocols in Internet of Things (IOT)," *Int. J. Comput. Appl.*, vol. 160, no. 2, pp. 18–22, 2017, doi: 10.5120/ijca2017912973.

[17] L. Atzori, A. Iera, and G. Morabito, "SIoT: Giving a Social Structure to the Internet of Things," *IEEE Commun. Lett.*, vol. 15, no. 11, pp. 1193–1195, 2011, doi: 10.1109/LCOMM.2011.090911.111340.

[18] V. Caballero, D. Vernet, and A. Zaballos, "Social Internet of Energy - A New Paradigm for Demand Side Management," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9853–9867, 2019, doi: 10.1109/JIOT.2019.2932508.

[19] M. Nitti, R. Girau, and L. Atzori, "Trustworthiness Management in the Social Internet of Things," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1253–1266, 2013, doi: 10.1007/s11277-017-4319-8.

[20] Z. Lin and L. Dong, "Clarifying Trust in Social Internet of Things," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 2, pp. 234–248, 2018, doi: 10.1109/TKDE.2017.2762678.

[21] M. A. Azad, S. Bag, F. Hao, and A. Shalaginov, "Decentralized Self-Enforcing Trust Management System for Social Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2690–2703, 2020, doi: 10.1109/JIOT.2019.2962282.

[22] U. Bodkhe, D. Mehta, S. Tanwar, P. Bhattacharya, P. K. Singh, and W. C. Hong, "A survey on decentralized consensus mechanisms for cyber physical systems," *IEEE Access*, vol. 8, pp. 54371–54401, 2020, doi: 10.1109/ACCESS.2020.2981415.

[23] Y. Fang, P. Chen, G. Cai, F. C. M. Lau, S. C. Liew, and G. Han, "Outage-limit-approaching channel coding for future wireless communications: Root-protograph low-density parity-check codes," *IEEE Veh. Technol. Mag.*, vol. 14, no. 2, pp. 85–93, 2019, doi: 10.1109/MVT.2019.2903343.

[24] J. Gaelens, P. Van Torre, J. Verhaevert, and H. Rogier, "Lora mobile-to-base-station channel characterization in the Antarctic," *Sensors*, vol. 17, no. 8, p. 1903, 2017, doi: 10.3390/s17081903.

[25] X. Pan, F. Di Maio, and E. Zio, "A benchmark of dynamic reliability methods for probabilistic safety assessment," in *2017 2nd International Conference on System Reliability and Safety (ICSRS)*, 2017, pp. 82–90, doi: 10.1109/ICSRS.2017.8272801.

[26] M. Castro and B. Liskov, "Practical Byzantine Fault Tolerance and Proactive Recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, 2002, doi: 10.1145/571637.571640.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

SISCOM: Smart Services for Information Systems and Communication Networks

Mónica Aguilar Igartua, Luis Javier de la Cruz Llopis, Jordi Forné Muñoz, Esteve Pallarès Segarra, Francisco José Rico Novella

Departmento de Ingeniería Telemática. Universitat Politècnica de Catalunya (UPC)

monica.aguilar@upc.edu, luis.delacruz@upc.edu, jordi.forne@upc.edu, esteve.pallares@upc.edu, francisco.jose.rico@upc.edu

The SISCOM (Smart Services for Information Systems and Communication Networks) research group focuses on technologies that make it possible to provide intelligent services for information services and communication networks. We teach and do research in topics related to privacy, performance evaluation of networks, wireless adhoc and mesh networks, design of routing protocols, among others. Our research activities are funded by public research projects granted by the Spanish Government and the European Commission.

***Keywords*- Communication networks, security and privacy of the information**

I. Introduction

The SISCOM (Smart Services for Information Systems and Communication Networks) research group [1] focuses on technologies that make it possible to provide intelligent services for information services and communication networks. More specifically, the interests of the group are mainly focused on two areas:

1. Communication networks: We work mostly on wireless networks, either with infrastructure or adhoc and mesh. With the aim of improving their performance, telematics engineering techniques are used. Also, routing protocols are designed and evaluated to guarantee the quality of service required to support intelligent services.

2. Security and privacy of the information: We work on the development of information protection techniques, especially on anonymization of databases to protect users' privacy when these data are being analysed by third parties. These techniques allow us to protect communications and personal information in data analysis processes.

Application examples: Anonymization of medical databases, protection of user privacy in Web browsing, electronic voting, communication in wireless networks with/without infrastructure (vehicular networks, mesh networks, mobile adhoc networks), design of efficient routing protocols for wireless infrastructureless networks, personal networks, machile learning algorithms (autonomous vehicles, routing protocols, quality of service).

Social impact: Health sector, administration, electric operators, smart grid, telecommunications service operators, smart city services, electrical vehicle, autonomous vehicle.

II. Research Projects

A. The UPC team has a deep expertise in data privacy, covering areas such as design of privacy mechanisms and metrics, anonymization algorithms and differential privacy. The members have participated in several national and European projects, and carried out contracts with companies such as NEC Labs and Microsoft. The team has a large experience in the field of multihop wireless networks, including congestion control mechanisms for smart grid neighborhood area networks, some of them based on machine learning techniques. Also, the team has a long experience on the design of QoS-aware multimetric routing protocols for vehicular networks in urban scenarios, some of them using ML-based models.

We also highlight our most recent projects:

MAGOS (2018-2020): Secure sMArt Grid using Open Source Intelligence. TEC2017-84197-C4-1/2/3-R
INRISCO (2015-2019): INcident monitoRing In Smart COmmunities. TEC2014-54335-C4-1/2/3/4-R.

Acknowledgements

This work was supported by the Spanish Government under research projects "sMArt Grid using Open Source intelligence (MAGOS)" TEC 2017-84197-C4-3-R and "Enhancing Communication Protocols with Machine Learning while Protecting Sensitive Data (COMPROMISE)" PID2020-113795RB-C31/AEI/10.13039/501100011033.

References

[1] Grupo de Investigación "Smart Services for Information Systems and Communication Networks (SISCOM)", https://siscom.upc.edu


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Optimizing the Response Time in SDN-Fog Environments for Time-Strict IoT Applications

Juan Luis Herrera, Jaime Galán-Jiménez, Javier Berrocal, Juan M. Murillo
Departamento de Ingeniería de Sistenas Informáticos y Telemáticos
Universidad de Extremadura
Avda. de la Universidad, S/N, Cáceres, España
{jlherrerag, jaime, jberolm, juanmamu}@unex.es

The Internet of Things (IoT) paradigm has brought to applications the potential of automating real-world processes. Applying IoT to intensive domains comes with strict Quality of Service (QoS) requirements. To achieve such goals, a first option is to distribute the computational workload throughout the infrastructure (edge, fog, cloud). In addition, its integration with Software-Defined Networks (SDN) can even further improve the QoS experienced, thanks to the global network view of the SDN controller. Therefore, the best placement for both the computation elements and the SDN controllers must be identified to optimize QoS. To obtain a truly optimal result, it is crucial to solve the problem in a single effort. In this work, a framework to identify the optimal deployment for distributed applications, DADO, is proposed, implemented and evaluated over an IIoT case study. DADO achieves response times up to 15.42% shorter than state-of-the-art benchmarks.

***Palabras Clave*—Fog computing, Internet of Things, Software-Defined Network**

I. SUMMARY

We live surrounded by everyday *things* that are connected to the Internet and run IoT applications – programs that interact with the real world. This interaction has generated interest in intensive domains such as industry or healthcare. However, integrating IoT applications into these domains is complex, as they have very strict QoS requirements. For this reason, cloud computing is very complicated to leverage, because it imposes a large latency penalty. New paradigms, such as fog computing, propose bringing some of the computing resources closer to the network edge, hence enhancing the QoS. Nonetheless, to properly make use of fog paradigms, it is important to optimally distribute the application's microservices among the available resources. This distribution problem is known as the Decentralized Computation Distribution Problem (DCDP). It is also important to note that the QoS of the application is determined not only by the computing QoS, but also by the networking QoS. Within the networking dimension, SDNs are also key enablers for fog paradigms because of the flexibility, scalability, and network programmability provided by the SDN paradigm, that allow for common tasks in IoT applications, such as service discovery, to be performed transparently. The QoS of SDNs, however, heavily depends on the QoS between the network's switches and the SDN controller they are assigned to. The problem in which controllers are placed and assigned optimally to SDN switches is the SDN Controller Placement Problem (CPP).

Both problems are deeply related, as the decisions on microservice placement are related to the networking QoS, and the CPP is heavily affected by the steering of application traffic flows. Thus, both problems require to be solved jointly in order to obtain optimal QoS. In this paper, we present Distributed Application Deployment Optimization (DADO), a framework based on mixed integer linear programming that jointly solves the DCDP and the CPP to minimize the response time in SDN-Fog environments. An evaluation of DADO over an industrial case study shows that this framework provides scalable solutions with response times up to 37.89% lower than alternative solutions, and up to 15.42% shorter than state-of-the-art benchmarks.

In the future, we expect to extend DADO, developing heuristics that will allow DADO to be applied to infrastructures larger than 300 nodes and while finding near-optimal solutions, as well as to add mobility considerations to DADO. Finally, we intend to expand DADO to consider multiple QoS features, such as reliability.

This work has been published in IEEE Internet of Things Journal, Vol 8, Issue 11, 2021. Impact Factor: 9.936. DOI: https://doi.org/10.1109/JIOT.2021.3077992. This work has been partially funded by the project RTI2018-094591-B-I00 (MCI/AEI/FEDER,UE), the 4IE+ Project (0499-4IE-PLUS-4-E) funded by the Interreg V-A España-Portugal (POCTEP) 2014-2020 program, by the Department of Economy, Science and Digital Agenda of the Government of Extremadura (GR18112, IB18030), by the Valhondo Calaff institution and by the European Regional Development Fund.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Teoría de grafos e inteligencia colectiva para análisis de opinión a gran escala

Marino Tejedor Romero[12], David Orden Martín[2], Encarnación Fernández Fernández[3], Iván Marsá Maestre[1], José Manuel Giménez Guzmán[1], Luis Cruz Piris[1]
[1]Departamento de Automática, Universidad de Alcalá
[2]Departamento de Física y Matemáticas, Universidad de Alcalá
[3]Departamento de Ingeniería Agrícola y Forestal, Universidad de Valladolid
{marino.tejedor,david.orden,ivan.marsa,josem.gimenez,luis.cruz}@uah.es; encarnacion.fernandez@uva.es

Que las redes telemáticas han proporcionado un entorno fértil para el análisis masivo de la opinión colectiva y de las preferencias individuales es algo fuera de toda duda. Compañías de todo el mundo recopilan datos sobre los usuarios para monetizarlos en forma de perfiles publicitarios, estudios de mercado, sociológicos, o políticos. Entre las herramientas de análisis de grandes cantidades de datos que hoy resultan más prometedoras está la teoría de grafos, por su potencial para inferir relaciones sutiles entre conceptos. Este potencial se está intentando explotar tanto desde el ámbito académico [1] como empresarial (e.g. *Graphext* en España o *HiveWise Inc.* en USA). En general, los grafos se emplean para modelar el conocimiento a partir de los datos.

En este caso, lo que proponemos es usar grafos también para obtener las opiniones de los usuarios. Para ello pretendemos evolucionar SensoGraph, una herramienta basada en grafos que se ha demostrado muy útil para el análisis sensorial en tecnología de alimentos. Se basa en el mapeo proyectivo, que consiste en que el usuario posicione los elementos sobre los que manifiesta su opinión en un plano de forma que más cercanía implique más similaridad, y viceversa. Una vez recabada la información de diferentes usuarios, se emplean grafos de proximidad para la agregación y la extracción de conclusiones [2]. Con esto se consigue una mejora importante en la eficiencia computacional con el número de participantes. Esto abre la puerta a la realización de análisis con una participación masiva, como puede verse en la figura.

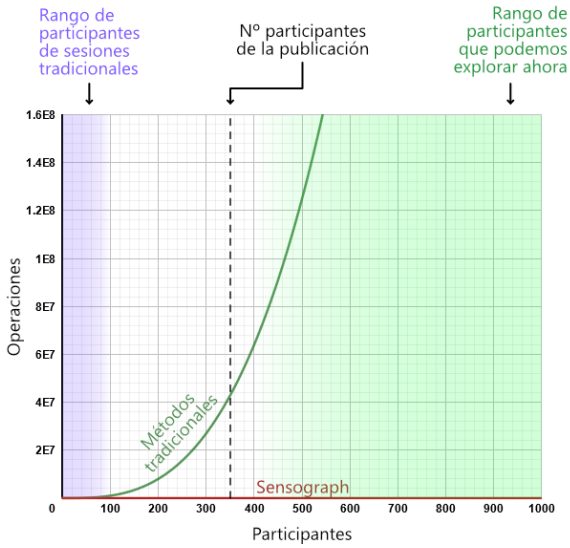
En esta contribución discutimos la aplicación de SensoGraph a análisis sensorial con un número elevado de participantes. Se ha utilizado una aplicación Web desarrollada por los autores que automatiza la entrada de datos y el procesamiento para recoger la opinión de 349 participantes, un número hasta entonces nunca visto en análisis sensorial, gracias a la optimización de las técnicas de análisis empleadas, así como a la eficiencia en su implementación.

Ahora estamos trabajando en extender estos métodos de recolección y agregación a escenarios de análisis masivo de opinión. Estamos haciendo experimentos en entornos diversos, desde la tecnología de alimentos [2] al análisis sobre el fraude académico. También estamos trabajando en análisis de opinión sobre productos a través de Internet, basadas en imágenes de productos y sus empaquetados.

***Palabras Clave*- teoría de grafos, inteligencia colectiva**

AGRADECIMIENTOS

Project UCeNet (CM/JIN/2019-031) of the Comunidad de Madrid and University of Alcalá



REFERENCIAS

[1] Klein, Mark. "Crowd-Scale Deliberation for Group Decision-Making." *Handbook of Group Decision and Negotiation* (2021): 355.

[2] Orden, D., Fernández-Fernández, E., Tejedor-Romero, M., & Martínez-Moraian, A. (2021). Geometric and statistical techniques for projective mapping of chocolate chip cookies with a large number of consumers. *Food Quality and Preference*, *87*, 104068

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Presentación de "Attention to Wi-Fi Diversity: Resource Management in WLANs with Heterogeneous APs".

José Saldana, José Ruiz-Mas, Julián Fernández-Navajas, José Luis Salazar.
Departamento de Ingeniería Electrónica y Comunicaciones.
Universidad de Zaragoza - Calle María de Luna, 3, 50018 Zaragoza (Spain).
jmsaldana@fcirce.es, jruiz@unizar.es, navajas@unizar.es, jsalazar@unizar.es

I. RESUMEN DEL TRABAJO YA PUBLICADO

El presente trabajo se ha centrado en escenarios de Wi-Fi que integran una pequeña cantidad de AP (normalmente de 2 a 4) con características heterogéneas, incluidos los últimos avances de 802.11n y 11ac, y opciones de seguridad. En él, se ha propuesto considerar las capacidades específicas y otras características para mejorar la gestión de los recursos radioeléctricos en la red.

Además, se ha construido una nueva aplicación de espacio de usuario, que trabaja en coordinación con un controlador existente, para construir una solución de Software Defined Wireless Network (SDWN) basada en LVAP. Esto permite transferencias fluidas y permite una gestión inteligente de las estaciones terminales (STA). La aplicación también gestiona la seguridad, evitando la necesidad de agregar elementos específicos para la autenticación, y el cifrado y descifrado.

Mediante tres casos de uso, se ha demostrado que la consideración de las diferentes características de los Access Point (AP) es relevante y puede aportar importantes beneficios para este tipo de soluciones, como también en cualquier WLAN coordinada con AP heterogéneos.

Se ha ampliado el protocolo existente para la comunicación entre el AP y el controlador SDWN, con el fin de comunicar y almacenar las características específicas de cada AP y STA, incluidas las capacidades, las características de seguridad y el estado. Esto permite el uso de estos parámetros mediante nuevos algoritmos de gestión de recursos que se pueden desarrollar en el futuro. Se ha realizado una batería de pruebas utilizando diferentes equipos, mostrando que los traspasos entre bandas de frecuencia son posibles, a la vez que se ha estimado los retrasos de procesamiento, el RTT y el retraso de traspaso, que es lo suficientemente pequeño para no producir ninguna interrupción significativa al usuario.

Finalmente, los escenarios considerados se han replicado en un entorno de simulación. Los resultados cuantitativos obtenidos muestran que se pueden lograr beneficios significativos si se consideran las características específicas de cada AP.

Como trabajo futuro, se pueden utilizar herramientas de simulación para explorar y medir los beneficios alcanzables mediante la consideración de las características heterogéneas de los AP (derivadas, en muchos casos, de la configuración particular del AP), con el fin de asignar cada STA conectado al AP que mejor se adapte a sus características. En estos entornos se pueden probar algoritmos inteligentes para la optimización de recursos. Además, la solución propuesta podría compararse (cualitativa y cuantitativamente, y también en términos de seguridad) con EasyMesh de Wi-Fi Alliance.

Palabras Clave- 802.11, seamless handoff, Software Defined Wireless Network (SDWN), Wireless LAN.

II. REFERENCIA DE LA PUBLICACIÓN ORIGINAL

J. Saldana, J. Ruiz-Mas, J. Fernandez-Navajas, J. Salazar, J. Javaudin, J. Bonnamy and M. Le Dizes, "Attention to Wi-Fi Diversity: Resource Management in WLANs With Heterogeneous APs," in IEEE Access, vol. 9, pp. 6961-6980, 2021,
doi: 10.1109/ACCESS.2021.3049180.

This original work has been financed by ORANGE as an External Research Contract "Wireless LAN based on use of Light Virtual WiFi Access Points," and partially financed by European Social Fund and Government of Aragon, CeNIT Research Group T31_20R.

AGRADECIMIENTOS

Este presentación de trabajo ha sido parcialmente financiado por el proyecto T31_20R del Gobierno de Aragón y RED2018- 102383-T del Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Grupo SMIoT: Sistemas Multimedia e IoT

Xabiel G. Pañeda, Roberto García, David Melendi, Laura Pozueco, Victor Corcoba, Sara Paiva, Dan García, Próspero Morán.
Departamento de Informática,
Universidad de Oviedo
Campus de Viesques, Gijón, Asturias
{xabiel, garciaroberto, melendi, pozuecolaura, corcobavictor, garciadan, moranprospero}@uniovi.es

El grupo de investigación SMIoT (Sistemas Multimedia e Internet de las Cosas) es un equipo multidisciplinar cuyos integrantes pertenecen, mayoritariamente, al área de Ingeniería Telemática de la Universidad de Oviedo, además de contar con médicos y especialistas en comunicación. La actividad del grupo en los próximos años se centrará en las líneas de investigación en envejecimiento activo y necesidades especiales y la línea de asistencia a la conducción de vehículos.

I. Envejecimiento Activo y Necesidades Especiales

Esta línea está centrada en la creación de un sistema cuyo interfaz de usuario sea un asistente de voz con pantalla con el objetivo de ayudar a personas con necesidades especiales (mayores y con enfermedades de diferentes tipos) a tener una vida más confortable.

Se creará un sistema de gestión para una red de dispositivos AMAZON ALEXA capaz de registrar usuarios indicando sus necesidades y patologías.

Se definirá un entorno de interacción mediante voces e imágenes de familiares. Para aumentar la cercanía y personalizar el método de interacción con el dispositivo asistente, se producirán avatares generados de familiares del receptor y fotos simplificadas.

Se creará un motor basado en IA capaz de tomar decisiones para recomendar actividades, estudiar el estado emocional de la persona, etc. Debe proponer actividades a sus usuarios y tomar decisiones sobre las actividades a proponer en base al contexto, las patologías del usuario y el historial de propuestas anteriores.

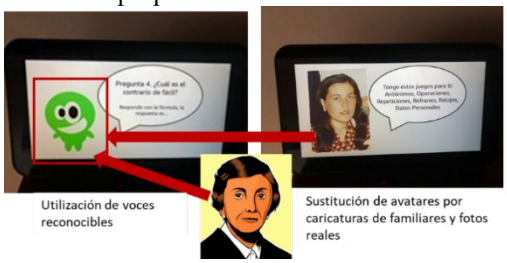


Fig. 1. Envejecimiento activo y necesidades especiales

II. Asistencia a la Conducción de Vehículos

La línea de investigación en asistencia a la conducción de vehículos está centrada en asistir a los conductores de diferentes tipos de vehículos de mejorar su eficiencia y seguridad. Para ello se monitorizarán diferentes tipos de elementos, se conectará con servicios de Internet, utilizando arquitecturas de software complejas como *Edge-Computing* se analizará la información, y se generarán diferentes tipos de avisos a los conductores. Todo ello se realizará buscando la mayor eficiencia y el mínimo impacto cognitivo sobre el conductor.

Se evaluarán diferentes métodos de interacción con el conductor ante la presencia de peatones en la zona de muerte para los sistemas de apoyo a vehículos industriales. En esta actividad se realizarán test con usuarios reales para evaluar la efectividad de diferentes tipos de avisos.

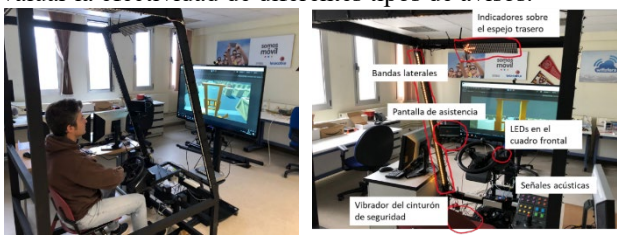


Fig. 2. Prototipo de asistencia a la conducción en vehículo industrial

Se evaluarán diferentes métodos de triangulación basados en *bluetooth* para la determinación de la localización del peatón en zona de riesgo.

Se construirá un simulador de conducción de carretillas elevadoras basado en el motor UNITY. Para llevar a cabo evaluaciones con usuarios es relevante que estos realicen una tarea lo más cercana al contexto en el que se desarrollaría la actividad real. Se creará un simulador capaz de interaccionar con el sistema de asistencia a la conducción con algunas características especiales:

- Aviso de obstáculos ocultos.
- Referencia al destino al que debe dirigirse el vehículo.
- Control de las señales a activar en el asistente

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Línea de Investigación en Ciberseguridad. Grupo TIC154 - Ingeniería Telemática

Rafael Estepa Alonso, Jesús Esteban Díaz Verdejo
Grupo TIC154-Ingeniería Telemática - Junta de Andalucía
Universidad de Granada, Universidad de Sevilla
rafaestepa@us.es, jedv@ugr.es

Presentamos las principales líneas de investigación del grupo TIC154-Ingeniería Telemática (del catálogo de grupos de investigación de Andalucía) relacionadas con la ciberseguridad, compuesto por investigadores de las Universidades de Sevilla y Granada.

***Palabras Clave*- grupo de investigación, ciberseguridad**

I. Introducción

El grupo TIC154–Ingeniería Telemática está compuesto por 13 investigadores [1], de los cuales 10 son doctores. El grupo es multidisciplinar, abarcando campos como el diseño de redes (NGN, VoIP, UAV), IoT (gestión flotas, optimización en WSN) o diseño de aplicaciones distribuidas en el contexto de la docencia o salud. En el seno de este grupo existe una activa línea de investigación en ciberseguridad, donde trabajan 6 miembros del grupo: 5 procedentes del Dpto. de Ingeniería Telemática de la Universidad de Sevilla y 1 del Dpto. de Teoría de la Señal, Telemática y Comunicaciones de la Univ. de Granada. Esta línea se encuentra actualmente en expansión, contando con proyectos activos y resultados de investigación y transferencia que serán descritos a continuación [2].

II. Líneas de Trabajo

El objeto de investigación es el desarrollo de sistemas y técnicas para proporcionar seguridad a servidores web y a los entornos industriales (ICS -tanto Industria 4.0 como IoT-). Para ello incorporamos los 4 elementos indicados en la Fig. 1 y que se describen a continuación:

Detección de intrusiones y anomalías. Podemos establecer dos bloques:

Fig. 1. Líneas de investigación.

- la detección de intrusiones en el servicio web en base a anomalías y la generación automatizada de firmas para SIDS, para lo que aplicamos técnicas basadas en modelado de Markov;
- la detección de anomalías en entornos industriales e IoT, en base la monitorización y clasificación de flujos a partir del análisis de una matriz de tráfico así como de anomalías del servicio o aplicación.

Modelado y clasificación de tráfico. Se desarrollan modelos de tráfico/comportamiento a nivel de flujos que, junto con clasificadores de flujos, se incorporan en el proceso de detección de anomalías.

Gestión de riesgos. Sistemas automatizados de análisis y priorización de riesgos en base a detectores de eventos basados en el análisis del tráfico y vulnerabilidades.

Acceso y Autenticación. Esta línea se centra actualmente en el control de acceso normalizado a datos sanitarios con monitorización de actividad por *Blockchain*.

III. Técnicas y Métodos

Las líneas de trabajo anteriores se apoyan en el análisis y modelado de protocolos de comunicaciones, técnicas de aprendizaje automático y minería de datos, modelado de procesos mediante modelos de Markov, análisis de grandes volúmenes de datos, despliegue y modelado de servicios.

IV. Resultados

Se han generado en los últimos años 22 publicaciones en revista y 17 en congresos, habiendo desarrollado 9 proyectos/contratos de investigación, la mayoría en estrecha colaboración con empresas del sector.

Referencias

[1] Disponible en: https://departamento.us.es/ingtelematica/#research . Último acceso 21 jun 2021.
[2] Disponible en: https://investigacion.us.es/sisius/sis_depgrupos.php?ct=&cs=&seltext=TIC-154&selfield=Cod

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Invirtiendo las Clases del Área de Ingeniería Telemática… poco a poco

Miguel A. Martín-Tardío, Jaime Galán-Jiménez.
Departamento Ingeniería de Sistemas Informáticos y Telemáticos
Universidad de Extremadura (UEx)
Avda. de la Universidad, s/n, 10003 Cáceres.
matardio@unex.es, jaime@unex.es.

Este trabajo presenta los resultados de la experiencia de la aplicación de una metodología de clase invertida en dos asignaturas impartidas por el Área de Ingeniería Telemática durante el curso 2020-21. Para ello, partimos de la base de numerosos estudios que muestran los beneficios del uso de *Flipped Classroom* en la docencia universitaria, como una participación más dinámica y participativa por parte de los estudiantes y un mejor aprovechamiento del tiempo. El análisis de los resultados académicos demuestra que se obtienen mejores resultados académicos. Además, los estudiantes muestran una opinión favorable de esta metodología respecto de la metodología tradicional, con afirmaciones relevantes como que perciben un mayor grado de aprendizaje de los conocimientos y que las interacciones con el profesorado son más frecuentes y positivas.

***Palabras Clave*- Clase invertida, flipped classroom, docencia en telemática, fundamentos de redes, redes de ordenadores**

I. Introducción

La idea general de la metodología *Flipped Classroom* [1] consiste en que lo que tradicionalmente se hacía dentro del aula (transmitir la información) ahora se hace fuera de clase a través de herramientas y recursos on-line. Así, los estudiantes dedican el tiempo en casa previo a la siguiente sesión síncrona (presencial o virtual) a estudiar los contenidos mediante actividades de aprendizaje de orden inferior (leer, ver vídeos). Por el contrario, las "tareas para casa" ahora se hacen en el tiempo de clase con la ayuda del profesorado y de los compañeros, realizando actividades cognitivas de orden superior (resolver problemas, estudiar casos prácticos) [2].

Cuando usamos el término *Flipped Classroom (FC)* debemos tener en cuenta que muchos modelos similares de instrucción se han desarrollado bajo otras denominaciones. Es el caso de *Peer Instruction* desarrollado por el profesor Eric Mazur en Harvard [3], que incorpora una técnica denominada *Just-in-Time Teaching (JiTT)* como un elemento complementario al modelo de clase invertida. JiTT permite al profesorado recibir retroalimentación de los estudiantes antes de la clase, para que éste pueda preparar estrategias y actividades para el aula centradas en las deficiencias detectadas en la comprensión del contenido [4]. Y otra metodología activa también complementaria a *FC* es *Team Based Learning (TBL)*. *TBL* se ha mostrado efectiva para mejorar el aprendizaje con trabajo colaborativo en equipo empleando la evaluación formativa [5]. Como los alumnos estudian cada tema o parte antes de la clase, la idea es proporcionar una estimulación para que realicen ese estudio previo ante la inminencia de una prueba de evaluación. Después en clase se realiza un cuestionario breve de evaluación con preguntas *Multiple Choice Questionnaries (MCQ)* que no evalúan recuerdo sino comprensión, primero de forma individual y más tarde en equipo, discutiendo con sus compañeros y recibiendo *feedback* sobre los errores entre ellos mismos y también del profesorado (evaluación formativa).

El objetivo de este trabajo es presentar los resultados de la aplicación de una metodología de clase invertida híbrida (*FC+JiTT+TBL)* en dos asignaturas del Área de Ingeniería Telemática de la UEx; centrándonos en la percepción de los estudiantes respecto al esfuerzo, la satisfacción general y los resultados académicos.

Este artículo se ha organizado en cinco apartados. Después de la introducción, se revisan brevemente otros trabajos previos sobre la clase invertida en la educación superior. Seguidamente se explica el método empleado y el contexto en el que se ha realizado la investigación. A continuación, se muestran los resultados del análisis realizado a partir de los datos obtenidos con las herramientas del apartado anterior y una discusión de los mismo, para terminar con las conclusiones.

II. Antecedentes

En el contexto de las enseñanzas universitarias en general, y de las ingenierías en particular, encontramos

numerosas revisiones que evalúan la aplicación de estas metodologías activas de forma general, pero no específicas en al ámbito de las Telecomunicaciones o la Informática [6]–[14]. En casi todas ellas, aparecen resultados consistentes sobre el efecto positivo en la percepción de los estudiantes sobre la satisfacción general con estas metodologías y una mejora del rendimiento académico.

Sin embargo, la implantación de estas metodologías no está exenta de dificultades. Por un lado, requiere un cambio de dinámica importante a los estudiantes acostumbrados a la clase magistral, lo que les supone un mayor esfuerzo. Y por otro, al profesorado por la necesidad de diseñar una tarea previa que proporcione contenido relevante y que debe percibirse como realizable; y unos cuestionarios de evaluación que procuren la discusión de los conceptos. Además, debe mantenerse la motivación de los estudiantes para la realización de la tarea previa a lo largo del tiempo con una despenalización evidente del error y la concesión de pequeñas recompensas con peso en la calificación final de la asignatura (p.e., insignias de cumplimiento) [2], [15], [16].

III. Metodología

Esta investigación se ha llevado a cabo para los últimos cuatro cursos de la asignatura Fundamentos de Redes (FR) del grado en Ingeniería Telemática en Telecomunicación impartida en el Centro Universitario de Mérida. Y por primera vez, durante el curso 2020-21, para la asignatura Redes de Ordenadores (RO) del grado en Ingeniería Informática en Ingeniería de Computadores impartida en la Escuela Politécnica de Cáceres. La estrategia para el seguimiento y mejora de la implantación de este modelo de clase invertida se centra en recoger las opiniones y resultados de los estudiantes en relación con:

- *El esfuerzo o carga de trabajo del estudio previo:* para investigar este asunto se ha decidido recoger como evidencia una declaración de las horas dedicadas al estudio previo de los contenidos antes de la/s clase/s semanal/es. Esto se realiza a través de la pregunta (CEE5) del cuestionario de estudio previo [17].
- *La satisfacción con el aprendizaje:* para recoger la percepción que los estudiantes tienen sobre la metodología y cómo influye en su aprendizaje de los contenidos. Se utiliza un cuestionario de elaboración propia de 16 preguntas [18]. La pregunta 1 (de respuesta Sí/No) y las preguntas 2-15 (con una escala Likert de 5 niveles "Muy en desacuerdo – Muy de acuerdo") son obligatorias. La pregunta 16 es una pregunta abierta opcional. Este cuestionario esta diseñado con preguntas que invitan a una práctica de reflexión y pensamiento crítico respecto a la formación recibida. Las encuestas de satisfacción han sido realizadas al final del período lectivo, antes de realizar la segunda prueba de evaluación sumativa, de forma anónima, obligatoria y presencial (comentar que la participación en las encuestas para la asignatura RO se vio afectada por la coyuntura provocada por la pandemia de COVID-19).
- *Los resultados académicos:* con la idea de investigar cómo haya podido influir en el rendimiento académico de los estudiantes, se ha decidido considerar las siguientes evidencias:
  - Número de NO presentados.
  - Calificaciones de la modalidad evaluación continua.

IV. Contexto

Las dos asignaturas constan de 60 horas de clase y 90 de estudio no presencial y son impartidas por diferentes profesores (Tabla I). Con respecto a la evaluación, un 40-50% de la nota final corresponde a la realización de dos exámenes de evaluación continua a lo largo del semestre, y el otro % a otras actividades de evaluación.

La idea es desarrollar y experimentar con un modelo de clase invertida híbrida similar a [19]–[21] que pueda implantarse de forma común en las asignaturas del Área de Ingeniería Telemática de la UEx. Y los puntos más relevantes del mismo son:

- Los contenidos de los temas están dispuestos semanalmente en una guía de estudio en la plataforma *Moodle* del campus virtual (para su estudio previo), acompañada de vídeos explicativos realizados por los profesores que deben ser vistos previamente a la clase (FC). Los vídeos, con una duración no superior a los 8-10 min., incorporan preguntas intercaladas para evaluación formativa y seguimiento de la visualización. Para ello se han empleado las herramientas *EdPuzzle* y *H5P* [22], [23].
- Después, los estudiantes deben completar un cuestionario *on-line* de comprobación del estudio previo (CCE) [17] a través de la plataforma *Moodle* como *feedback*, que los profesores utilizamos para diseñar la siguiente clase (JiTT).
- Ya en la clase, durante la primera parte, los estudiantes realizan un cuestionario MCQ de evaluación formativa sobre los contenidos revisados. Y después de obtener *feedback* a través las actividades programadas, vuelven a realizarlo al final de la clase, pero respondiendo en equipo (TBL). Ambos cuestionarios se realizan con la herramienta *Socrative* [24].
- Durante la parte central de la clase se realizan presentaciones cortas y; actividades prácticas, ejercicios, problemas y/o estudios de casos en grupo para trabajar las dificultades de aprendizaje indicadas a través del CCE.
- También se entregan pequeñas recompensas (insignias) en forma de puntuación para la evaluación continua por el cumplimiento de las actividades del estudio previo.

Tabla I
Asignaturas del AIT con modelo de aprendizaje invertido

Asignatura	Acrónimo	Semestre	Estudiantes
Redes de Ordenadores	RO	5	48
Fundamentos de Redes	FR	1	30

V. Resultados

Una vez identificados los indicadores de seguimiento, se exponen los resultados obtenidos tras la aplicación de la metodología de aprendizaje invertido seguida.

A. Experiencia previa

Aproximadamente el 70% de los alumnos participantes en estas asignaturas no han tenido una experiencia de aprendizaje anterior con este tipo de metodologías activas. Bien porque son estudiantes de nueva incorporación que ponen de manifiesto la escasa penetración de estas metodologías activas en los centros de secundaria de origen; bien porque no es un tipo de metodología extendida entre el profesorado de estas titulaciones.

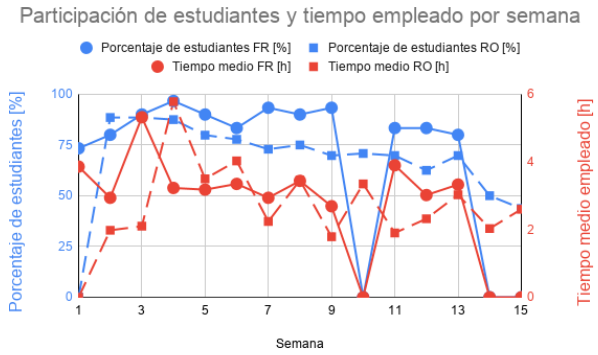


Fig. 1. Porcentaje de estudiantes y tiempo medio empleado en el estudio previo por semana.

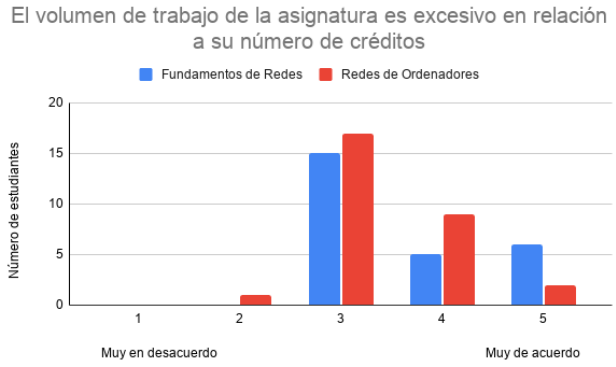


Fig. 2. Valoración del volumen de trabajo relacionado con el estudio previo en relación con el número de créditos.

B. Esfuerzo de los estudiantes

Como se muestra en la Fig. 1 los estudiantes comienzan muy motivados y con una alta participación (90%) que va decayendo según avanza el semestre. Sin embargo, esa caída no está relacionada directamente con un incremento en el tiempo medio de dedicación semanal. Un motivo plausible es que seguramente la carga de trabajo del curso en general aumenta al final del semestre, en concreto durante las dos últimas semanas. También puede observarse que en las semanas 10 y 15 no hubo estudio previo porque se realizaron actividades de evaluación sumativa. Por otro lado, según la pregunta 14 del cuestionario de satisfacción (Fig. 2), el 56,6% de los estudiantes percibieron un esfuerzo medio en relación con el número de créditos, y entre un 38-40% percibieron que el esfuerzo fue alto/muy alto. Sin embargo, el porcentaje medio de horas totales no presenciales de dedicación al estudio previo y la preparación de los dos exámenes de esas asignaturas ha sido aproximadamente del 45% de las horas no presenciales incluidas en los planes docentes.

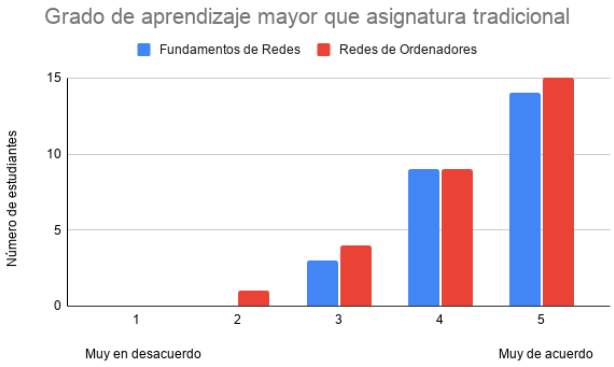


Fig. 3. ¿Tu grado de aprendizaje de los contenidos es mayor que una asignatura tradicional?

C. Satisfacción con el aprendizaje

En primer lugar, debe tenerse en cuenta que los datos obtenidos tienen limitaciones en cuanto al número de respuestas en las encuestas de satisfacción, con una participación del 71% de los estudiantes matriculados en ambas asignaturas (representan el 90% de FR y un 60,4% de RO).

Según los resultados obtenidos en el cuestionario de satisfacción, el 83,5% de los estudiantes de estas asignaturas perciben un mayor grado de aprendizaje de los contenidos (Fig. 3) y el 94% están satisfechos en general (pregunta 15) con esta propuesta metodológica, teniendo en cuenta que el 70% de ellos nunca había recibido docencia de esta manera. En general, puede constatarse esa satisfacción a partir de las reflexiones recogidas (pregunta 16) en el sentido de: está muy bien; se aprende muchísimo mejor; permite seguir la asignatura de manera más fácil; y a los estudiantes de primer año ayuda a entablar relaciones con personas nuevas, y a crear un clima de confianza con la clase y el profesor. Y entre las reflexiones para la mejora más destacables encontramos:

- "Hay ocasiones que tenemos demasiada carga de trabajo ya que se nos juntan exámenes de otras asignaturas y no le podemos dedicar tanto tiempo a preparar la clase de esta forma".
- "La duración de los vídeos ya que si son demasiado largos pierdes un poco el hilo (deberían durar entre 10-15 minutos)".
- "Al no estar acostumbrado a tener que trabajar el material antes de que el profesor lo explicase, me ha resultado algo pesado hacer los estudios

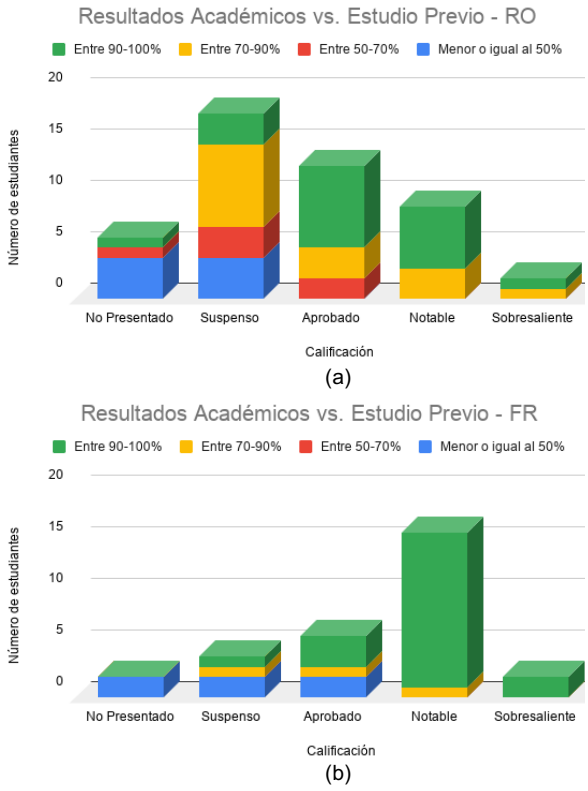


Fig. 4. Resultados académicos comparados con el cumplimiento del estudio previo realizado.

previos. Pero el planteamiento me parece correcto".

- "Me parece interesantísimo incorporar más videos y reducir la carga de lectura en la teoría, a mi me cuesta muchísimo sentarme a leer y estudiar, sin embargo, esto me facilita muchísimo el aprendizaje, me resulta mucho más interactivo".

De forma complementaria, también se realizó un análisis de correlación lineal a partir de los resultados del cuestionario de satisfacción [18]. Considerando como variable dependiente (Y) la pregunta 15 (*En general, estás satisfecho con la experiencia educativa planteada en esta asignatura*), como variables independientes X_1 (pregunta 2) hasta X_{13} (pregunta 14) y un intervalo de confianza del 95%, se obtuvo un coeficiente de determinación $R^2 = 0{,}65$ con un p-valor < 0,001. Así, puede afirmarse que el 65% de la satisfacción general del estudiante podría explicarse a partir de estas variables independientes, pero sólo X_4 (pregunta 5: *Tu grado de aprendizaje de los contenidos de los temas es mayor que en una asignatura tradicional*) con p-valor < 0,001 y X_6 (pregunta 7: *Hace que las interacciones con el/la profesora sean más frecuentes y positivas*) con un p-valor = 0,006 se mostraron como significativas en este análisis.

D. *Resultados académicos*

El análisis de los resultados académicos se ha realizado comparando las calificaciones obtenidas con el cumplimiento del estudio previo. La Fig. 4 muestra la relación entre las calificaciones obtenidas por los estudiantes (eje X) durante la convocatoria de enero y, para cada calificación, el nº de estudiantes (eje Y) según el porcentaje de realización del estudio previo (colores). Según la Fig. 4.a, el 50% de los estudiantes de RO han superado la asignatura, y de ellos más del 90% realizó estudio previo de los contenidos en algún momento. Sin embargo, es relevante entre los estudiantes suspensos, que la mitad de ellos también completasen el mismo porcentaje de estudio previo que los aprobados, pero no les valiese para superar esta asignatura. Teniendo en cuenta que la segunda coincide con la convocatoria de examen final, podría deberse a un problema de carga de trabajo. Aunque estos resultados son similares a los del curso anterior (no *Flipped Classroom)*, si se segregan las calificaciones obtenidas se descubre que el número de Notables aumentó de un 10% al 23%. Y esto podría relacionarse directamente con una mejor satisfacción del estudiante con el aprendizaje.

Según la Fig. 4.b, el 82,8% de los estudiantes de FR han superado la asignatura de los cuales también más del 90% realizó estudio previo. Este porcentaje de aprobados es mayor al 64,8% de media de aprobados desde el curso 2016-17, y no parece que exista una relación con un mayor cumplimiento del estudio previo, que para el curso pasado fue similar (96,4%). Lo que sí pone de manifiesto es que realizar menos del 50% del estudio previo proporciona una baja probabilidad de aprobar esta asignatura. Hay que comentar que en esta asignatura también se realizan dos pruebas de evaluación sumativa durante el semestre, pero a diferencia de RO, el segundo examen se realiza antes de la convocatoria del examen final. Además, durante el examen final existe la posibilidad de recuperar una de las dos partes pendiente si fuese el caso. En este curso, de los 11 estudiantes presentados al examen final, los que tenían sólo una parte pendiente (4) consiguieron aprobarla.

Por último, la incidencia de los No Presentados entre las dos asignaturas está en torno al 9% de los estudiantes matriculados, porcentaje similar al de cursos anteriores. Por tanto, no parece que podamos encontrar una relación directa, ni de mejora, ni empeoramiento de este porcentaje con el tipo de metodología docente aplicada.

VI. Conclusiones

Los resultados anteriores vienen a confirmar lo que la literatura sobre la clase invertida en el entorno universitario dice, que los estudiantes valoran muy favorablemente la utilización de esta metodología como una forma más ágil, dinámica e interactiva de adquirir los conocimientos. También que trabajan más los contenidos de las asignaturas, pero en su mayoría no lo asumen como una carga de trabajo excesiva respecto al número de créditos, y en general, obtienen mejores resultados académicos. Sin embargo, ¿qué ocurriría si esta metodología se aplicase de forma generalizada en una titulación donde no existe una masa crítica de profesores aplicándola con antelación?

Además, se mantienen motivados prácticamente a lo largo del semestre con un grado de participación y cumplimiento elevado del estudio previo a. En general, la satisfacción del estudiante aumenta frente a una asignatura con una metodología de clase magistral, y en particular, en el grado de aprendizaje de los contenidos que perciben. Por tanto, un adecuado diseño, planificación e implementación de las actividades de clase y los materiales de estudio semanales es fundamental para lograr ese éxito entre los estudiantes [25]. Y también destaca en esa satisfacción que

las interacciones profesorado-estudiantes sean más frecuentes y positivas. El desarrollo de una mayor cantidad de actividades de carácter práctico para el aula en detrimento de la actividad expositiva, aumenta la comunicación con los estudiantes por el propio proceso de aprendizaje y fomenta la discusión en el aula por la naturaleza de esas actividades. Así, los estudiantes estarían dispuestos a que otras asignaturas incorporasen este tipo de metodología a sus clases.

Por parte de los profesores, estos manifiestan que la preparación de los vídeos, la elaboración de más actividades para el aula y el seguimiento del estudio previo semanal para las asignaturas supone un incremento importante de trabajo. Este aumento de la carga de trabajo es innato a la adopción de esta metodología. No obstante, la carga de trabajo se estabiliza según avanzan los cursos por la reutilización de los materiales de cursos anteriores, la experiencia en el diseño semanal de las clases y el desarrollo de la habilidad de "automatizar" las actividades de seguimiento con apoyo de herramientas software como el campus virtual. Este profesorado en *Flipped Classroom* manifiesta una satisfacción muy alta con su propio desempeño docente relativa a aspectos como el aprovechamiento del tiempo de clase, la interacción con sus estudiantes, la participación y motivación de los estudiantes en el aula, y los resultados académicos.

Finalmente, vamos a resumir una serie de lecciones aprendidas y propuestas de mejora, tanto para el proceso enseñanza-aprendizaje, como para el proceso de evaluación de la docencia a partir de la experiencia acumulada:

- Trasladar la instrucción a casa a través de vídeos cortos no superiores a 8-10 minutos facilita el aprendizaje de los contenidos.
- Repensar los procedimientos de evaluación continua. La programación de actividades de evaluación sumativa a lo largo del curso, con la posibilidad de recuperar una parte durante el examen final, puede ayudar a los estudiantes a superar la asignatura reduciendo la "sobrecarga" del final de curso.
- Concentrar, si es posible, las sesiones de grupo grande (en nuestro caso 2 sesiones de 50 min.) en un sólo día para disponer de 100 minutos de clase continuada. Esto permite al profesor concentrar el estudio previo a lo largo de la semana para una única sesión y diseñar sesiones presenciales más completas. Esto permitiría incluir en todas las sesiones actividades *TBL* de evaluación formativa.
- Conceder pequeñas recompensas por el cumplimiento, tanto de las actividades de estudio previo, como las realizadas en el aula, mejora la motivación de los estudiantes. Esta puede ser una excelente manera de medir realmente el criterio "Participación en la asignatura", asignándole una puntuación de la nota final.
- Revisar la encuesta de satisfacción. Concretamente, revisar los enunciados de las preguntas para facilitar responder con asertividad. Por otro lado, las preguntas 3 y 14 pueden provocar confusión y deberían resumirse en una sola. Además, deberían incluirse nuevos ítems significativos para obtener la opinión del estudiante sobre aspectos como el uso del material en formato de vídeo o la motivación, e investigar cómo pueden influir en la percepción de la satisfacción general.
- Relacionado con lo anterior respecto al material en formato de vídeo, revisar las respuestas correspondientes a la pregunta CCE4 del cuestionario de comprobación del estudio previo (CCE) referida a la mejora de los materiales docentes, para incorporarla al análisis de los resultados.
- Finalmente, incorporar al proceso de evaluación de la docencia de final del semestre una encuesta SEEQ (*Student´s Evaluation of Educational Quality*) para disponer de evidencias que nos permitan revisar de forma permanente del cometido docente.

REFERENCIAS

[1] A. Sams and J. Bergmann, "Flip your students' learning," *Educ. Leadersh.*, vol. 70, no. 6, pp. 16–20, Mar. 2013.

[2] M. Marqués Andrés, "Qué hay detrás de la clase al revés (flipped classroom)," in *JENUI 2016*, 2016, pp. 77–84, Accessed: Feb. 08, 2020. [Online]. Available: http://schools.khanacademy.org/.

[3] E. Mazur, "Peer instruction: Getting students to think in class," in *AIP Conference Proceedings*, 1997, vol. 399, pp. 981–988, doi: 10.1063/1.53199.

[4] E. Mazur and J. Watkins, "Just-in-Time Teaching and Peer Instruction," *Just Time Teach. Across Discip. Across Acad.*, pp. 39–62, 2009.

[5] L. K. Michaelsen, A. B. Knight, and L. D. Fink, *Team-based learning: A transformative use of small groups*. Westport, Conn: Praeger, 2002.

[6] B. Kerr, "The flipped classroom in engineering education: A survey of the research," in *Proceedings of 2015 International Conference on Interactive Collaborative Learning, ICL 2015*, 2015, pp. 815–818, doi: 10.1109/ICL.2015.7318133.

[7] A. Prieto, J. Barbarroja Escudero, A. Corell, and S. Álvarez Álvarez, "Eficacia del modelo de aula invertida (flipped classroom) en la enseñanza universitaria : una síntesis de las mejores evidencias," *Rev. Educ.*, 2021, doi: 10.4438/1988-592X-RE-2021-391-476.

[8] J. L. Bishop and M. A. Verleger, "The flipped classroom: A survey of the research," 2013, doi: 10.18260/1-2--22585.

[9] J. Polastre, R. Szewczyk, A. Mainwaring, D. Culler, and J. Anderson, "Anaysis of Wireless sensor Networks for Habitat Monitoring," in *Wireless Sensor Networks*, C. S. Raghavendra, K. M. Sivalingam, and T. Znati, Eds. Norwell, MA, USA: Kluwer Academic Publishers, 2004, pp. 399–423.

[10] J. O'Flaherty and C. Phillips, "The use of flipped classrooms in higher education: A scoping review," *Internet High. Educ.*, vol. 25, pp. 85–95, Apr. 2015, doi: 10.1016/j.iheduc.2015.02.002.

[11] F. J. Hinojo-Lucena, I. Aznar, J. M. Romero, and J. A. Marín, "Influencia del aula invertida en el rendimiento académico. Una revisión sistemática," *Campus Virtuales*, vol. 8, no. 1, pp. 9–18, 2019.

[12] M. A. Al Mamun, M. A. K. Azad, M. A. Al Mamun, and M.

Boyle, "Review of flipped learning in engineering education: Scientific mapping and research horizon," *Educ. Inf. Technol.*, no. 0123456789, 2021, doi: 10.1007/s10639-021-10630-z.

[13] B. Prevalla and H. Uzunboylu, "Flipped learning in engineering education," *TEM J.*, vol. 8, no. 2, pp. 656–661, 2019, doi: 10.18421/TEM82-46.

[14] H. Al-Samarraie, A. Shamsuddin, and A. I. Alzahrani, *A flipped classroom model in higher education: a review of the evidence across disciplines*, vol. 68, no. 3. Springer US, 2020.

[15] A. Prieto Martín, "Flipped Learning. Aplicar el modelo de Aprendizaje Inverso," in *Narcea*, vol. 1, no. 1, 2017, pp. 109–125.

[16] S. Catalán and M. Marqués, "Aprendiendo bases de datos... al revés," in *JENUI 2018*, 2018, pp. 13–19.

[17] J. Galán-Jiménez and M. A. Martín-Tardío, "Cuestionario de comprobación del estudio previo (CCE)," *Google Drive*, 2020. http://bit.ly/3oYZd1X.

[18] J. Galán-Jiménez and M. A. Martín-Tardío, "Encuesta de satisfacción de la metodología invertida," *Google Drive*, 2020. http://bit.ly/2N7IJqS.

[19] A. Prieto Martín *et al.*, "Nuevas combinaciones de aula inversa con just in time teaching y análisis de respuestas de los alumnos," *RIED. Rev. Iberoam. Educ. a Distancia*, vol. 21, no. 1, p. 175, Sep. 2017, doi: 10.5944/ried.21.1.18836.

[20] N. Lasry, M. Dugdale, and E. Charles, "Just in Time to Flip Your Classroom," *Phys. Teach.*, vol. 52, no. 1, pp. 34–37, Jan. 2014, doi: 10.1119/1.4849151.

[21] N. Rowley and J. Green, "Just-in-time Teaching and Peer Instruction in the Flipped Classroom to Enhance Student Learning," *Educ. Pract.*, vol. 2, no. 1, pp. 14–17, 2015.

[22] "EdPuzzle," 2021. https://edpuzzle.com (accessed Jun. 01, 2021).

[23] "H5P," 2021. https://h5p.org (accessed Jun. 01, 2021).

[24] "Socrative," 2021. https://www.socrative.com (accessed Jun. 01, 2021).

[25] M. Marqués and J. M. Badía, "¿Qué nos dicen los estudiantes sobre lo que hace que funcione la clase invertida?," *JENUI 2021*, vol. 6, pp. 51–58, 2021.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Estrategias de fomento del trabajo continuo en modalidades semipresenciales

Guillermo Azuara, Julián Fernández-Navajas, José María Saldaña, José Luis Salazar, José Ruiz-Mas, Antonio Valdovinos, José García, María Ángela Hernández, María Canales, José Ramón Gállego, Álvaro Alesanco, Ignacio Martínez.
Departamento de Ingeniería Electrónica y Comunicaciones
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza
{gazuara, navajas, jsaldana, jsalazar, jruiz, toni, jogarmo, anhersol, mcanales, jrgalleg, alesanco, imr}@unizar.es

Resumen.- **En este artículo se presenta el conjunto de estrategias utilizadas durante los dos últimos cursos académicos en las asignaturas del área de Ingeniería Telemática, así como su valoración cualitativa y su aplicación a las clases no presenciales impuesta por la situación sanitaria. En el curso 2019-2020 los profesores del área de Ingeniería Telemática de la Universidad de Zaragoza participaron en un proyecto de innovación docente para medir el impacto de las diferentes estrategias de fomento del trabajo continuo en diferentes asignaturas y titulaciones, aunque debido a la situación derivada por la pandemia esta comparación y evaluación de dicho impacto ya no tiene sentido por la gran cantidad de factores externos que han afectado estos dos últimos cursos a los alumnos, se presentan los resultados obtenidos de la aplicación de estas estrategias durante los mismos. No se han detectado especiales diferencias en la aplicación de las estrategias entre asignaturas. Además, los alumnos se adaptaron en general muy bien a la nueva situación y nuevas metodologías. El proyecto ha permitido conocer distintas experiencias y actividades, así como sus ventajas e inconveniente, lo que permitirá ampliar el abanico de herramientas disponibles para mejorar la docencia en el futuro.**

Palabras Clave- **docencia online, telemática, trabajo continuo, vídeos.**

I. Introducción

Durante los cursos 2019-2020 y 2020-2021 todos los profesores del área de Ingeniería Telemática participaron en el proyecto de innovación docente de la Universidad de Zaragoza "Estudio del impacto de diferentes estrategias para fomentar el estudio continuo de los estudiantes en la adquisición de competencias", cuyos objetivos eran: presentar diferentes estrategias para fomentar el trabajo diario en las asignaturas y la motivación; estudiar el impacto de las diferentes estrategias en los resultados académicos (que deberían medir el logro de las competencias previstas); comparar las peculiaridades de utilización de las mismas estrategias en diferentes asignaturas; y conocer y aprender de las experiencias del resto de integrantes del proyecto.

Los docentes participantes imparten clase en dos centros: la Escuela de Ingeniería y Arquitectura (EINA) y la Escuela Universitaria Politécnica de Teruel (EUPT). Se plantearon diferentes actividades previas a las actividades docentes evaluables (ya sea una prueba de evaluación de contenidos o una práctica). La valoración de las actividades podía ser: no cuenta para la nota de la asignatura, no cuenta para la nota de la asignatura pero está gamificada o sí que contará para la nota final. El planteamiento inicial era estimar el grado de implicación de los alumnos en la estrategia mediante la nota obtenida en la actividad (aunque no cuente para la evaluación final, sí se valorará lo que los alumnos hayan entregado) y comparar la calificación final media de la actividad evaluable (realizada después) con la calificación media de cursos pasados, para poder comparar la eficacia de las diferentes estrategias, siendo conscientes de las diferencias intrínsecas que puede haber entre los grupos de alumnos de cursos distintos. También se diseñó un cuestionario a los alumnos y otro al profesor para que valoren el esfuerzo en la actividad y sus percepciones.

Se creó un repositorio común donde se iban subiendo los diferentes informes de las actividades, para que todos los integrantes del proyecto pudieran conocer el trabajo y las reflexiones de los demás compañeros.

Con la irrupción de la pandemia en marzo de 2020 y sus posteriores consecuencias a nivel docente (reorganización de actividades, preparación de entornos de trabajo, clases on-line, etc…) no se pudieron realizar las tareas conforme estaba previsto (sólo había datos de una asignatura del primer semestre). El curso 2020-2021 también estuvo salpicado de limitaciones y alteraciones en el normal desarrollo de la docencia, por lo que se decidió seguir con la parte del proyecto de contar las diferentes acciones que se iban llevando en las asignaturas, pero sin

tener en cuenta los resultados cuantitativos. Adicionalmente, en muchas asignaturas se hicieron reflexiones relativas a buenas prácticas para la docencia virtual, problemas detectados y posibles soluciones.

En la sección II se presentará el contexto del proyecto: las asignaturas que finalmente se han analizado y una justificación (más allá de la intuición) de por qué creemos que los datos no son comparables con otros cursos. En la sección III se presentará la metodología de recogida de información que se planteó para la toma de datos y compartición de resultados y reflexiones. En la parte IV se presentarán las actividades realizadas y su valoración cuantitativa. En la sección V se presentarán los resultados obtenidos y finalmente en la VI se presentarán las conclusiones.

II. Contexto del proyecto

A lo largo de diversas experiencias docentes realizadas en cursos previos, una carencia de cara a plasmar los resultados en publicaciones académicas es la necesidad de resultados objetivos que orienten sobre las mejoras (o no) alcanzadas con diversas estrategias. En este proyecto se pretendía comparar los resultados obtenidos tras la aplicación de diferentes estrategias y metodologías, con los obtenidos en cursos previos sin realizar estas actividades, para poder comparar los resultados. Como ya se ha comentado, dada la situación sanitaria de los últimos cursos, se ha centrado en resultados cualitativos ya que, como se verá a continuación, no es posible comparar con cursos anteriores por las condiciones de contorno radicalmente distintas. Para más información sobre análisis a nivel mundial de impactos, respuestas políticas y recomendaciones ver [1], y para tener una visión de medidas adoptadas a nivel general en la universidad española ver [2].

Realizando una búsqueda en cualquier base de datos bibliográfica de los términos "*COVID teach university engineering*" nos sorprenderemos (o no) de la gran cantidad de artículos que se han publicado desde 2020 (más de 22.000 aparecen en Google Scholar, habría que ver en detalle los que realmente versen sobre estos temas, o en cualquier caso siendo algo más rigurosos más de 130 en *Scopus* sin incluir en la búsqueda el término "*engineering*", en cuyo caso sólo aparecen 13 pero podríamos localizar más utilizando en la búsqueda otros términos relacionados como communications, networks, etc…). Realizando una revisión rápida y no exhaustiva de los artículos de la rama de ingeniería, podemos comprobar que a lo largo y ancho del mundo la respuesta de la educación superior ha sido bastante similar (un poco improvisada al principio y ya más elaborada en el último curso): clases on-line síncronas, vídeos grabados con explicaciones, uso de simuladores para la realización de prácticas siempre que ha sido posible (o grupos reducidos si no era posible simularlas) y plataformas de formación (como Moodle) como eje de la comunicación docente-estudiantes. Algo que también hemos visto en el análisis de las asignaturas que hemos trabajado en el proyecto.

Hay muchos artículos que señalan el importante impacto sicológico de la situación vivida en los estudiantes [3 - 8]. En nuestro caso, podemos tener identificados a los estudiantes que han estado infectados o han tenido que guardar cuarentena por contacto estrecho, pero muchos otros han sufrido pérdidas familiares u hospitalizaciones de familiares con el estrés físico y sicológico que esto acarrea, y estos datos no los tenemos disponibles, pero dados los porcentajes de afectados en las diversas olas en Aragón, es muy probable que tuviéramos un número importante de estudiantes afectados por estas circunstancias, por ello se decidió no comparar las calificaciones con los cursos previos y quedarnos en valoraciones cualitativas, ya sea a partir de encuestas normalizadas o a partir de otras formas de recibir información (conversaciones informales con estudiantes principalmente).

La elección de las asignaturas que se han analizado finalmente en el proyecto obedece a la decisión de los docentes participantes, teniendo en cuenta las circunstancias excepcionales de impartición de la docencia en estos dos últimos cursos, el desarrollo concreto de las asignaturas y la información que se ha podido obtener de cada una. En la elección no se han tenido en cuenta otros factores más allá de intentar tener asignaturas de dos centros y diferentes cursos y titulaciones. Las asignaturas estudiadas son sólo una pequeña parte de las que imparte el área, y han sido las siguientes:

- Redes de Computadores (EUPT): Grado de Ingeniería Informática, 2º curso primer semestre.
- Diseño y Administración de Redes (EINA y Escuela Universitaria Politécnica de Teruel): Grado de Ingeniería Informática, 4º curso primer semestre.
- Fundamentos de Redes (EINA): Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, 1er curso, segundo semestre.
- Interconexión de redes (EINA): Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, 2º curso, segundo semestre.
- Programación de redes y servicios (EINA): Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, 2º curso, primer semestre.
- Análisis y Dimensionado de Redes (EINA): Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, 3er curso, primer semestre.

III. Metodología de recopilación de información

Al comenzar el proyecto se creó una carpeta compartida con Google Drive a la que tenían acceso todos los miembros del proyecto.

Se preparó un texto con las instrucciones para la recogida de datos y se creó una hoja de cálculo con Google Sheets para homogeneizar la recogida de datos en todas las asignaturas. Los datos que se solicitaban eran los siguientes (se creó una hoja para cada asignatura): Asignatura, actividad realizada, número total de alumnos en la asignatura y número de alumnos participantes en la actividad (relevante si ésta es voluntaria). Además, tras la realización evaluable que se realizaba tras la actividad había que completar una pequeña tabla con dos filas, una para los alumnos que habían realizado la actividad y otra para los que no, en la que había que indicar para cada conjunto la calificación mínima del grupo, el cuartil 1, el cuartil 3, la calificación máxima, la mediana y la media. A

partir de estos datos se creaba un gráfico de velas que permitía una comparación rápida visual entre los resultados de ambos grupos. Notar que, con este sistema, al proporcionar sólo estos valores y al no estar identificados los estudiantes, no existen problemas de protección de datos.

Además de estos datos esquemáticos, para cada asignatura se debía rellenar un pequeño informe con los siguientes campos:

- Datos de la asignatura: nombre, código, curso, semestre, titulación, centro, número de alumnos matriculados y profesores que han participado en la actividad:
- Datos de la actividad: número de alumnos que han realizado la actividad, tipo de actividad (Test / Preguntas abiertas / Vídeo / Otras) y descripción detallada de la actividad (siendo este campo donde se realiza la explicación más detallada de la actividad propuesta, obligatoria para los alumnos (sí/no) y si cuenta para nota detallar el peso.

Finalmente, para la valoración se diseñaron dos cuestionarios de satisfacción de la actividad, uno para los alumnos y otro para los docentes. Las preguntas para los alumnos eran preguntas típicas de escala Likert de cinco niveles junto a otras de texto libre o respuesta numérica:

- ¿Le ha parecido interesante la actividad? (elegir valor de 1-5).
- ¿Cuánto tiempo ha dedicado aproximadamente a preparar cada actividad (en minutos)? (introducir valor).
- ¿Cree que gracias a la actividad ha obtenido un mejor resultado? (elegir valor de 1-5).
- ¿Le gustaría tener más actividades de este tipo? (elegir valor de 1-5).
- Si desea hacer algún comentario positivo use este campo (texto libre).
- Si desea hacer alguna crítica use este campo (texto libre).
- Si desea hacer alguna sugerencia use este campo (texto libre).

Las preguntas para los profesores eran:

- ¿Cuánto tiempo ha dedicado a preparar la actividad (en minutos)? (introducir valor).
- ¿Cree que ha merecido la pena el trabajo invertido en la actividad? (elegir valor de 1-5).
- ¿Cree que gracias a la actividad los alumnos han aprendido más o mejor? (elegir valor de 1-5).
- Si desea hacer algún comentario positivo use este campo (texto libre).
- Si desea hacer alguna crítica use este campo (texto libre).
- Si desea hacer alguna sugerencia use este campo (texto libre).

Como ya se comentando, esta planificación inicial se vio alterada por la necesidad imperiosa de reorganizar la docencia a los nuevos escenarios impuestos, y al final no fue posible obtener los informes de todas las asignaturas, no obstante, sí que se pudieron recabar opiniones valiosas de docentes y alumnos, aunque en un número mucho menor al inicialmente planteado.

IV. Actividades realizadas y valoración.

A continuación, se van a describir brevemente las principales actividades que se propusieron:

Prácticas 365 (novedoso): es una experiencia integrada de enseñanza-aprendizaje diseñada para las sesiones prácticas de laboratorio con un enfoque innovador, sostenible a lo largo de otros cursos y transferible a otras materias o disciplinas. Todo esto es posible gracias a que las innovaciones propuestas se han creado con Moodle y PowerPoint, estos recursos interactivos y audiovisuales "conectan" con los estudiantes y se han generado más de 25 indicadores para cuantificar las mejoras obtenidas "antes, durante y después". Esta experiencia se denomina Prácticas 365 porque incluye:

un aprendizaje previo a cada sesión de prácticas para responder a "3" cuestiones Moodle sin límite de intentos y con realimentación, en las que los estudiantes desarrollan el autoaprendizaje (cada cuestionario previo proporciona el resultado de cada una de las 3 cuestiones Moodle planteadas, con estos resultados, al empezar cada sesión, el profesorado conoce la preparación que cada estudiante ha realizado de la sesión confirmando si ha relacionado correctamente los fundamentos teóricos con los conceptos prácticos que va a ser necesario aplicar en el laboratorio); un trabajo práctico de laboratorio, impartido durante 10 sesiones de 2 horas cada una, y distribuido académicamente en "6" bloques didácticos en un entorno semipresencial, personalizado y gamificado (gracias a los cuestionarios previos, el profesorado puede dar feedback al estudiante durante la sesión práctica de forma personalizada, además, durante cada una de las sesiones prácticas, se realiza un seguimiento proactivo de cada estudiante mediante una secuencia de preguntas y respuestas que se puntúan con tarjetas de colores y estas valoraciones complementan la evaluación global como indicadores); y una evaluación final de "5" cuestiones Moodle para valorar todo el aprendizaje en progresión continua (con estos resultados, el profesorado evalúa progresivamente el nivel de aprendizaje intermedio de cada estudiante pudiendo detectar en tiempo real el aprendizaje a reforzar en las siguientes sesiones y los aspectos concretos que se recomiendan trabajar durante el curso, además todos estos indicadores de progresión continua se complementan con metodologías semipresenciales de seguimiento y aprendizaje como tutorías personalizadas, foros online, videollamadas, etc.).

Vídeos explicaciones docentes (tradicional): los docentes dejan accesibles videos (fueron creados el curso 2019-2020) con explicaciones sobre los temas a tratar en

la clase teórica, para ser visualizados antes de asistir a clase para mejorar el aprovechamiento. Este aprovechamiento se notó sobre todo en las sesiones de dudas de prácticas, que se aprovecharon mucho mejor. Por el contrario, parece que disponer de este material desincentiva la presencia física posterior en el aula, por lo que se trabajará en este sentido.

Trabajos prácticos ligados a las prácticas de laboratorio (se venía realizando en algunas de las asignaturas)**:** el objetivo no es incrementar el trabajo que debe realizar el alumno sino animar a que éste se desarrolle de forma sostenida a lo largo de todo el tiempo. Con esta motivación, el trabajo práctico, por ejemplo desarrollado alrededor de los despliegues de red que se simulan en el laboratorio, se introducen cuestiones teóricas y prácticas, planteadas de distintas formas, para verificar de forma progresiva la asimilación de los contenidos que se van desarrollando en las clases de teoría. Todo ello, haciendo un énfasis especial en la asimilación de los conceptos teóricos más relevantes. Estas cuestiones se suman a aquellas que constituyen el núcleo de las prácticas de laboratorio, pero permiten profundizar en aquellos aspectos que los alumnos deberían trabajar de forma autónoma como parte del estudio continuo de la asignatura. Con esta metodología se ha constatado que los alumnos más implicados consiguen un seguimiento más adecuado y una mejor realimentación sobre la asimilación de los conceptos que utilizan en el desarrollo de los modelos matemáticos.

Cuestiones cortas al final de cada tema: los alumnos contestan a las preguntas desde su móvil a un cuestionario de Moodle sobre el tema. No es obligatorio, pero la participación suele ser alta. Intenta que los alumnos lleven más al día el estudio de la asignatura. Más información en [9].

Cuestiones cortas antes de la práctica: tras haber contestado un guion de prácticas previo a la práctica (es obligado entregar pero no se evalúa) contestan a unas cuestiones cortas sobre el estudio previo y sí que cuentan para nota (ver [9]).

Videos cortos realizados por los alumnos: los alumnos en grupo trabajan un tema y deben tratar unos puntos que se les ha indicado en un vídeo de máximo 3 minutos. La actividad es voluntaria y en general la participación alta (este último año, debido a la pandemia la participación bajó). Se esfuerzan por hacer el video atractivo ya que luego es mostrado a los compañeros. En general da buenos resultados. Ver [10].

Para el desarrollo de estas metodologías se han utilizado principalmente herramientas de uso habitual: Moodle, Google Forms, MS-powerpoint, OBS Studio, MS-Excel, Google drive y Google Meet.

V. Resultados.

En la Tabla 1 podemos ver un resumen de las metodologías utilizadas en cada una de las asignaturas. Se utilizan las siguientes siglas: P365 (Prácticas 365), VED (Vídeos explicaciones docentes), TPL (Trabajos prácticos ligados a las prácticas de laboratorio), CCFT (Cuestiones cortas al final de cada tema), CCAP (Cuestiones cortas antes de la práctica), VCA (Videos cortos realizados por los alumnos).

Asignatura	P365	VED	TPL	CCFT	CCAP	VCA
Redes de Computadores			X	X	X	X
Diseño y Administración de Redes			X		X	
Fundamentos de Redes	X					
Interconexión de redes		X				
Programación de redes y servicios		X				
Análisis y Dimensionado de Redes			X			

Tabla 1. Metodologías por asignatura

A partir de las calificaciones (que no se han podido comparar con otros cursos anteriores por los motivos expuestos anteriormente), los informes de las actividades y las encuestas de satisfacción de estudiantes y profesores se presentan los resultados de las asignaturas analizadas.

Redes de Computadores: Grado de Ingeniería Informática (GII), 2º curso primer semestre. Docencia 100 % presencial. Se realizaron 11 cuestionarios, 5 tipo test o preguntas cortas al finalizar un tema (en clase con móvil, voluntario) y 6 obligatorios en las prácticas de laboratorio (2 tipo test previos, en clase, con el móvil y un peso del 25 % de cada práctica y 4 más extensos de seguimiento de prácticas a entregar a posteriori (75 % de cada práctica). La participación en los cuestionarios de tema fue descendiendo: porcentaje de participación: 69, 54, 58, 38 y 46). Las notas obtenidas en los cuestionarios han sido satisfactorias: media de cada cuestionario sobre 10: 5,33 – 7,88 – 6,85 – 7,13 – 8,56.

Se continuará trabajando la motivación y el diseño de los cuestionarios para que sean más motivadores y se mejore la participación.

Debido a las medidas sanitarias excepcionales motivadas por la pandemia no se pudo utilizar el laboratorio habitual, se tuvieron que rehacer los guiones de prácticas, y por tanto los resultados no son comparables con el año anterior.

En las encuestas específicas relacionadas con el proyecto los alumnos han valorado muy positivamente los cuestionarios (aunque se quejaron de la dificultad y el escaso tiempo, aspectos que ya se están trabajando). En una escala Linker (1 peor valor y 5 mejor valor): ¿Le ha parecido interesante la actividad? Media 4,5, mínimo 3 y máximo 5.

Los docentes valoran positivamente la actividad y aunque cuesta un esfuerzo extra creen que merece la pena.

Diseño y Administración de Redes: GII, 4º curso primer semestre. EINA on-line (incluye sesiones síncronas), EUPT 100 % presencial. Previo a cada práctica deben completar un cuestionario similar al que completarán después para anticipar los resultados. Obligatorio y no cuenta para nota. Al finalizar la práctica deben completar un cuestionario con los resultados obtenidos en la práctica. Obligatorio y que sí cuenta para nota.

Se considera que los cuestionarios son un estímulo importante al trabajo continuo del estudiante y a la preparación adecuada de las prácticas. Suponen un mejor aprovechamiento de las prácticas, lo que se puede medir con la reducción de tiempo en la realización de las mismas, lo que ha permitido realizar prácticas adicionales y resolución de problemas grupales.

Se sigue trabajando en la mejora de los cuestionarios.

Debido a las medidas sanitarias excepcionales motivadas por la pandemia no se pudo utilizar el laboratorio habitual, se tuvieron que rehacer los guiones de prácticas, y se trabajó con un entorno gráfico de simulación de redes (GNS3) (aprender su manejo también consume horas) por tanto los resultados no son comparables con el año anterior.

Resultados: aun teniendo en cuenta las excepcionales circunstancias, similares a los del curso anterior, lo que se considera positivo.

Los docentes valoran positivamente la actividad y aunque supone un extra de trabajo creen que da buenos resultados.

Fundamentos de Redes: Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación (GITST), 1er curso. La actuación se ha enfocado a las prácticas de laboratorio que se imparten en 10 sesiones (de 2 horas cada sesión) en 8 grupos (de unos 14 estudiantes/grupo). Para esta experiencia se ha desarrollado una metodología propia denominada prácticas 3-6-5: que consta de un aprendizaje previo a cada sesión de prácticas para responder a 3 cuestiones Moodle sin límite de intentos y con realimentación, en las que los estudiantes desarrollan el autoaprendizaje; Un trabajo práctico de laboratorio, impartido durante 10 sesiones de 2 horas cada una; Y una evaluación final de 5 cuestiones Moodle para valorar todo el aprendizaje en progresión continua

En las encuestas específicas del proyecto los alumnos han valorado muy positivamente la actividad. En una escala Linker (1 peor valor y 5 mejor valor): ¿Le ha parecido interesante la actividad? Media 4,55, mínimo 3 y máximo 5.

Los profesores también creen que del trabajo invertido para preparar las actividades se nota una mejora en la asimilación de conceptos por parte de los alumnos.

Interconexión de redes: GITST, 2º curso, segundo semestre. En este caso y debido a la pandemia además de impartición de clase tele-presenciales se prepararon vídeos explicativos docentes. Estos vídeos son una demanda generalizada por parte de los alumnos, pero desmotiva la participación en clase, por lo que a raíz de lo analizado en el proyecto se planteará un seguimiento de su visionado (mediante cuestionarios, debates relacionados, etc. siguiendo modalidades de docencia inversa) que fomentará una utilización más productiva de los mismos, ya que en la actualidad no se ha observado un mayor aprendizaje con su disponibilidad.

Programación de redes y servicios: GITST, 2º curso, primer semestre. Teoría tele-presencial, prácticas on-line. Como incentivo para una mejor preparación de la asignatura se ha aprovechado el material grabado del curso 2019-2020 (del periodo de confinamiento domiciliario), poniéndolo a disposición de los alumnos antes de las sesiones de teoría. En cualquier caso, si bien se les ha instado a los alumnos a su visionado previo, no se ha verificado la realización de la actividad, por lo que no se puede valorar el impacto real. En el caso de las prácticas, ha permitido al profesorado reducir el tiempo de explicaciones previas maximizando el aprovechamiento de las sesiones para la resolución de dudas, pero no hay una cuantificación del aprovechamiento real que hayan podido tener los alumnos.

Como mejora, se planteará un seguimiento del visionado del material audiovisual (mediante cuestionarios, debates relacionados, etc. siguiendo modalidades de docencia inversa) que fomentará una utilización más productiva de los mismos.

Análisis y Dimensionado de Redes: GITST, 3er curso, primer semestre. La asignatura, con una fuerte componente matemática, incluye la realización de 15 horas de prácticas de laboratorio, correspondientes a tres prácticas, donde los alumnos utilizan el software de simulación de redes OPNET para crear distintos despliegues de red, configurando distintos escenarios de generación de tráfico, de tal forma que el objetivo es obtener estadísticas de los parámetros de comportamiento más relevantes (retardo, utilización de recursos, *throughput*) y comparar los resultados con los obtenidos en modelos matemáticos que deben desarrollar a tal efecto, usando las herramientas aprendidas en la asignatura. El seguimiento de las clases está fuertemente condicionado por la asimilación de los conceptos introducidos en clases previas. La experiencia docente indicaba que se producía una postergación en el estudio de la asignatura que hipotecaba el aprendizaje. A pesar de programarse las clases prácticas y de requerir éstas el desarrollo de un estudio teórico, los alumnos no abordaban la asignatura hasta la fecha en la que estaba programada la práctica. Como consecuencia el aprovechamiento real era deficiente. Como medida correctora se ha planteado el desarrollo de trabajos prácticos ligados a las prácticas de laboratorio.

El objetivo era incentivar al alumno para que el estudio se desarrolle de forma sostenida a lo largo de todo el tiempo. Con esta motivación, el trabajo práctico, desarrollado alrededor de los despliegues de red que se simulan en el laboratorio, introduce cuestiones teóricas y prácticas, planteadas de distintas formas, para verificar de forma progresiva la asimilación de los contenidos que se van desarrollando en las clases de teoría. Todo ello, haciendo un énfasis especial en la asimilación de los conceptos teóricos más relevantes. Estas cuestiones se suman a aquellas que constituyen el núcleo de las prácticas de laboratorio, pero permiten profundizar en aquellos aspectos que los alumnos deberían trabajar de forma autónoma como parte del estudio continuo de la asignatura. Los trabajos prácticos se empiezan a desarrollar desde la segunda semana de impartición de la asignatura y la entrega definitiva se planifica con antelación al desarrollo de las prácticas.

Con esta metodología se ha constatado que los alumnos más implicados consiguen un seguimiento más adecuado

y una mejor realimentación sobre la asimilación de los conceptos que utilizan en el desarrollo de los modelos matemáticos. Mencionar que la valoración por parte de los alumnos de la asignatura (encuestas oficiales de la Universidad de Zaragoza) en su conjunto y de las preguntas 9 ("Metodología adecuada en relación con los objetivos de formación") y 10 ("Utilización de recursos didácticos (audiovisuales, de laboratorio, de campo, etc.)" en particular ha sido alta.

A nivel general se puede destacar que los alumnos se adaptaron en general muy bien a las condiciones, y aunque las relaciones con los docentes en la mayoría de las ocasiones eran por videoconferencia o correo electrónico, terminaron el curso satisfechos con la formación recibida y valorando el esfuerzo de los docentes. De las encuestas de satisfacción respecto a las actividades indicadas en este trabajo de alumnos y docentes se desprende que:

- Los alumnos valoran positivamente que "se les fuerce" (sic) a trabajar el material a lo largo de la asignatura (especialmente antes de una práctica).
- Notar que algunas de estas tareas, aunque no contaban para nota, eran obligatorias. Se entendía que era el trabajo normal de preparación que debía hacer el alumno para aprovechar correctamente la práctica y se solicitaba su entrega como motivación (muchos no lo hacían). No obstante, cuando la documentación que deben "cuenta para la nota" el trabajo es más completo y por lo general el resultado de la práctica más satisfactorio.
- En el caso de la experiencia de cuestionarios breves al final de cada tema de teoría, sorpresivamente se ha visto en los resultados finales no hay una correlación estrecha entre las calificaciones de los cuestionarios y los resultados finales. Probablemente pueda deberse a un tamaño demasiado reducido de la muestra.
- En general los alumnos más implicados con la asignatura valoran más este tipo de actividades. A la pregunta si les han parecido interesantes las actividades un 56,7 % otorgó la máxima puntuación de la escala (5) y un 40,3 % otorgó un 4. Sólo el 3 % otorgó un 3.
- El 92,5 % de los encuestados respondieron que creían que gracias a las actividades obtenían un mejor resultado (valoraciones de 4 o 5).
- Se ha valorado muy positivamente el potencial de este tipo de actividades para incorporarlas a nuevas partes de las asignaturas.

VI. Conclusiones.

La pandemia que nos ha asolado durante el último año y medio ha afectado de forma muy importante al proyecto de innovación docente inicialmente planteado, ya que uno de sus objetivos principales era cuantificar el impacto de diferentes estrategias para fomentar el aprendizaje continuo en la adquisición de competencias. El segundo cuatrimestre del curso 19-20 fue complicado y hubo que adaptar las metodologías docentes (especialmente las prácticas de laboratorio) para impartir las clases de forma remota. En el curso 20-21 se recuperó la presencialidad de forma parcial (EINA) o total (EUPT), aunque con una distorsión importante de las condiciones de docencia habituales: tamaño de grupos, imposibilidad de prácticas en grupos, limitaciones de aforos y acceso a laboratorios, etc. Aún a pesar de todo, se cree que sí se han podido sacar conclusiones interesantes del proyecto, al menos para poder hacer una valoración global. Como se ha indicado, los alumnos se adaptaron en general muy bien a la nueva situación y nuevas metodologías.

No se han detectado especiales diferencias en la aplicación de las estrategias entre asignaturas. El planteamiento en general ha sido muy similar, y en general las actividades propuestas funcionan de forma similar. Cuando no son obligatorias, son seguidas por los alumnos más implicados en las asignaturas.

El proyecto ha permitido conocer a todos los participantes distintas experiencias y actividades, sus ventajas e inconvenientes y, por tanto, de la experiencia se puede ampliar el abanico de herramientas disponibles para mejorar tanto la docencia en general como resolver algunos de los problemas concretos detectados.

Se considera que las diferentes actividades presentadas son fácilmente replicables en otras áreas y asignaturas, ya que se han utilizado herramientas de uso habitual (y muchas de ellas gratuitas) que se disponen en la mayoría de las universidades.

En el futuro se intentará dar solución a las carencias detectadas y llevar a cabo una recogida de datos amplia siguiendo las metodologías planteadas, para poder tener una valoración cuantitativa de los resultados.

Agradecimientos

Este trabajo ha sido financiado durante el curso 2019-2020 y 2020-2021 por el Programa de Incentivación de la Innovación Docente en la Universidad de Zaragoza (PIIDUZ_19_154).

Referencias

[1] Giannini, S., "Covid-19 y educación superior: De los efectos inmediatos al día después", Revista Latinoamericana de Educación Comparada, vol. 11, no 17, 2020, p. 1-57.

[2] "La Universidad frente a la pandemia. Actuaciones de Crue Universidades Españolas ante la COVID19". [Online]. Available: https://www.crue.org/wp-content/uploads/2020/12/La-Universidad-frente-a-la-Pandemia.pdf. [Accessed: 18-Jun-2021].

[3] Revilla-Cuesta, V., et al., "The Outbreak of the COVID-19 Pandemic and its Social Impact on Education: Were Engineering Teachers Ready to Teach Online?.", International Journal of Environmental Research and Public Health, 18.4,2127, 2021.

[4] de Oliveira Araújo, Francisco Jonathan, et al., "Impact of Sars-Cov-2 and its reverberation in global higher education and mental health.", Psychiatry Research ,288, 112977, 2020.

[5] Velázquez, Lilia González, "Estrés académico en estudiantes universitarios asociado a la pandemia por COVID-19.", Espacio I+D, Innovación Más Desarrollo, 9.25, 2020.

[6] Odriozola-González, P., Planchuelo-Gómez, A., Irurtia, M.J., de Luis-García, R., "Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university", Psychiatry Research, Volume 290,113108, 2020.

[7] Browning MHEM, Larson LR, Sharaievska I, Rigolon A, McAnirlin O, et al., "Psychological impacts from COVID-19 among university students: Risk factors across seven states in the United States", PLOS ONE 16(1): e0245327, 2021.

[8] Padrón, I., Fraga I. Vieitez L., Montes C., Romero E., "A Study on the Psychological Wound of COVID-19 in University Students", Frontiers in Psychology, 12, 2021.

[9] Azuara G., Fernández Navajas, J., Saldaña, J.M., Ayuso, N., Alastruey, J., "El Cuestionario como Recurso para la Mejora Docente", XIV Jornadas de Ingeniería telemática (JITEL 2019), Libro de actas, 2019, pp.: 204-208.

[10] Azuara Guillén, G., Fernández Iglesias, D., López Torres, AM., Salinas Baldellou, AM., Aguilar Martín, MC., Salazar Riaño, JL., Fernández-Navajas, J., "Vídeos cortos realizados por los alumnos como recurso docente. Diferentes enfoques.", en XIII Jornadas de Ingeniería telemática (JITEL 2017), Libro de actas, Editorial Universitat Politècnica de València, 2018, pp.: 348-355. https://doi.org/10.4995/JITEL2017.2017.6566

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Sistema telemático de citas para la docencia

Antonio Estepa, Antonio L. Delgado, Rafael Estepa
Departamento Ingeniería Telemática,
Universidad de Sevilla
C/Camino de los descubrimientos s/n.
{aestepa,aldelgado,rafaestepa}@us.es

En este trabajo presentamos un sistema de gestión de citas diseñado por el Dpto. de Ingeniería Telemática de la Universidad de Sevilla para satisfacer las necesidades de profesores y alumnos. Este servicio puede ser utilizado para concertar tutorías, seguimientos de proyectos o revisiones de exámenes. Durante un año de operación del sistema hemos podido probar sus efectos positivos en el desempeño docente y el aprendizaje de los alumnos. A los alumnos, les permite solicitar una cita en un horario que a ellos les sea más conveniente y tener la certeza de que el profesor estará esperándoles (presencialmente, o telemáticamente). También les permite reducir el tiempo de espera en las revisiones de exámenes. El profesorado puede saber con antelación cuándo vendrán los alumnos y qué tema desean tratar y así tener preparada la reunión, gestionando el tiempo de forma más eficaz. Además, le permite establecer los horarios disponibles para citas de forma flexible de manera que se adecúe a su disponibilidad dinámicamente (p.ej., semanalmente). También permite evitar aglomeraciones en las revisiones de exámenes. Finalmente, el sistema genera numerosos indicadores con los que el profesor puede gestionar el proceso de tutorías e integrarlo en una estrategia de mejora continua.

***Palabras Clave*- tutorías, cita previa, telemática**

I. INTRODUCCIÓN

Las sesiones de tutorías tienen una gran importancia para el aprendizaje del alumnado y existen estudios que así lo demuestran [1][2]. Quizás por ello, las universidades públicas españolas exigen a su profesorado una dedicación de 6 horas semanales (casi similar al encargo docente de un profesor) a la actividad tutorial. Además, fuera de estas horas regladas, el profesorado suele entrevistarse con los alumnos en reuniones de seguimiento de proyectos en las (cada vez más frecuentes) asignaturas de aprendizaje basado en proyectos [3], o revisiones de exámenes.

Lamentablemente, la experiencia docente de los autores muestra que el alumnado de su Escuela Superior de Ingeniería no suele aprovechar esta oportunidad de aprendizaje, siendo evidente que el sistema de tutorías actual no resulta atractivo para muchos estudiantes. El profesorado, mayoritariamente, suele ver con normalidad o indiferencia la falta de alumnos en tutoría y, en general, aprovecha dichas horas para realizar otras actividades ante la falta de asistencia. Los departamentos y la universidad no disponen de indicadores de uso de tutorías por parte de los alumnos u otro tipo de mecanismo de seguimiento, y delegan la gestión del servicio en cada profesor que, tampoco suele disponer de indicadores (p.ej., de asistencia) con lo que podríamos decir que su nivel de gestión actual es muy pobre. Sospechamos que esta situación no es exclusiva de la Escuela de Ingenieros, o de la Universidad de Sevilla, y que puede ser algo endémico en muchas otras universidades públicas españolas.

Una de las posibles causas del poco éxito de las tutorías podría ser la rigidez horaria del servicio. Siguiendo un paradigma quizás ya anacrónico, el profesorado de la Universidad de Sevilla tiene la obligación de publicar un horario fijo de tutoría que debe mantener a lo largo de todo el semestre. Cada profesor puede basar su horario en sus propias preferencias y no existe ninguna comprobación a posteriori sobre la compatibilidad o idoneidad de dicho horario con el horario de los alumnos a los que se pretende servir. Otra posible explicación del poco éxito de las tutorías es la exigencia de presencialidad para el alumnado. Tradicionalmente se ha exigido que las tutorías sean presenciales lo que dificulta que los alumnos, que ya han terminado las clases, vengan de nuevo al campus para una reunión (generalmente corta). Afortunadamente, desde el inicio de la actual pandemia, la necesidad de presencialidad se ha eliminado y el profesorado se ha adaptado mayoritariamente al uso de sistemas de teleconferencia (p.ej. *Microsoft teams*, *Google meet*, o los propios de los sistemas de enseñanza virtual de cada universidad como *Blackboard* o *Moodle*) para realizar tutorías telemáticas. En general estos sistemas se han probado exitosos y han sido bien acogidos por los alumnos, aunque es previsible que tras el fin de la pandemia una parte del profesorado vuelva a la exigencia de presencialidad. No obstante, aún con horarios más flexibles y tutorías telemáticas, el número de tutorías usadas por los alumnos durante la pandemia no parece haber subido notablemente, por lo que es necesaria una

investigación mas profunda sobre las causas de la falta de atractivo de las tutorías.

Aún con el sistema telemático, hay aspectos de gestión que afectan a la entrega del servicio y que podrían ser mejorados con la ayuda de sistemas de información. Dado que se presupone que el profesor esta disponible y en espera durante el horario oficial de tutoría, no existe un mecanismo de concertación de cita. Los beneficios de los sistemas de cita previa han sido estudiados ampliamente en el mundo de la sanidad [4], pero existen pocos estudios en el ámbito de la docencia universitaria. Parece conveniente disponer de un sistema de concertación de cita previa ya que evitaría al profesor el engorro de coordinarse por email con numerosos alumnos. Además, un sistema de cita previa evitaría al profesorado mantener una sesión abierta durante todo el horario oficial de tutoría (en el caso de las tutorías telemáticas). También permitiría al profesor conseguir información antes de la reunión (haciéndola más productiva). Además, un sistema de cita previa también sería útil para atender otros temas tales como reuniones de seguimiento de un proyecto, o revisiones de exámenes fuera del horario de tutoría "oficial". Otro problema no resuelto es el problema de la necesidad de establecer un horario predeterminado e inflexible para todo el semestre. Esto no sólo puede afectar negativamente al alumnado, sino que también puede resultar inconveniente para el profesorado cuando le surgen otras citas ineludibles que se desconocen en el momento de poner el horario (p.ej., un congreso, reunión de un proyecto, etc..). Sería ideal que el profesor supiera con antelación cuándo va a venir un alumno y qué desea tratar. También parece lógico que exista cierta flexibilidad en los horarios para poder así acomodar eventos inesperados o ineludibles y poder también adaptarse dinámicamente a las necesidades de los alumnos.

En este trabajo presentamos un sistema de gestión de citas que puede ser utilizado para tutorías, seguimientos o revisiones de exámenes. El sistema ha sido creado por el Dpto. de Ing. Telemática de la Universidad de Sevilla específicamente para satisfacer las necesidades de profesores y alumnos, y lleva en operación dos años. Durante este tiempo hemos podido probar sus efectos tanto antes de la pandemia como durante. La contribución del sistema desarrollado es notable para profesores y alumnos. Para los alumnos, les permite solicitar una cita en un horario que a ellos les sea más conveniente y tener la certeza de que el profesor estará esperándoles (presencialmente, o telemáticamente). También les permite recudir el tiempo de espera en las revisiones de exámenes. El profesorado puede saber con antelación cuándo vendrán los alumnos y qué tema desean tratar y así tener preparada la reunión y poder gestionar mejor su tiempo. Además, le permite la disposición de los turnos de cita de forma flexible de manera que se adecúe a su disponibilidad dinámicamente (p.ej., semanalmente). También permite evitar aglomeraciones en las revisiones de exámenes. Finalmente, el sistema genera numerosos indicadores con los que el profesor puede gestionar el proceso de tutorías e integrarlo en una estrategia de mejora continua.

El resto del artículo se estructura de la siguiente forma. En la sección II, ofrecemos un breve estado del arte sobre los sistemas de gestión de citas actualmente disponibles en el mercado. En la sección III se exponen los requisitos de diseño del aplicativo y su propuesta de valor. En las secciones IV y V se expone el diseño y la implementación de la aplicación. En la sección VI se muestra la operación de la aplicación y en la VII se describen algunos de los indicadores de gestión que genera la aplicación y que permiten la mejora del servicio. Finalmente, las secciones VIII y IX muestran algunas opiniones de los usuarios y conclusiones.

II. Sistemas de gestión de citas

Existen numerosos sistemas informáticos para la gestión de citas. En [5] y [6] podemos encontrar sendas comparativa entre sistemas de gestión de citas existentes. En la mayoría de los casos, los sistemas comparados son comerciales y están diseñados para satisfacer los objetivos de los procesos del negocio donde se aplican. Podemos clasificarlos en tres grandes grupos en función del ámbito de aplicación:

- Sistemas de gestión de citas en el ámbito sanitario. Estos sistemas, propuestos tanto académicamente [7] como comercialmente [8], tienen por objetivo la optimización del flujo de trabajo de las citas los sistemas sanitarios, mejorar la productividad de los trabajadores, gestionar los recursos de forma óptima y reducir el tiempo de espera de los usuarios. Estos sistemas suelen ofrecer funcionalidades propias del ámbito sanitario tales como reserva de múltiples citas relacionadas y con uso de múltiples recursos (i.e., pruebas, edificios, etc..). También suelen estar integrados en sistemas de información más complejos EHR (*Electronic Health Record*) que gestionan procesos, personas y recursos en este ámbito. Otros sistemas comerciales más modestos (p.ej., *doctoralia*, *Appointment Plus*, etc.) ofrecen el servicio de citas integrado con otros servicios (p.ej., video consulta) para clientes y consultas particulares, y suelen incluir recordatorios y gestión de otros aspectos útiles para las clínicas como el cobro por servicio.
- Sistemas de gestión de citas en el ámbito de la actividad empresarial. Estos sistemas comerciales tienen por objetivo mejorar la productividad y el servicio a los clientes. Suelen ser usados por negocios como salones de belleza, gimnasios, spam etc. Los más simples ofrecen sincronización con calendarios, citas automáticas, y gestión de pagos, los más complejos incluyen: citas de grupos, aplicaciones móviles, citas recurrentes, base de datos de clientes, e integración con los sistemas CRM de las empresas [9].

  Estos sistemas ofrecen múltiples canales de interactuación (email, web, redes sociales, chats) y cuidan la imagen que ofrecen a los clientes. Todos cobran dinero en base a una suscripción mensual cuyo importe varía en base a las funcionalidades incluidas. Suelen incluir la gestión de recursos y localizaciones de la cita y gestionan los pagos por las citas. Dos ejemplos de este tipo de sistemas son *Wix* o *Timify*.
- Sistemas de propósito genérico usables en la docencia. Algunos sistemas de calendario como Google Drive

Calendar o Microsoft One profesional pueden ser usados para la concertación de citas entre profesores y alumnos, pero el uso puede ser tedioso para la acción tutorial y obligan a la creación de usuarios "cautivos" en el sistema de Google o Microsoft, lo cual es intrusivo para los alumnos. Un sistema para concertar citas a veces empleado en la docencia es *Doodle*, cuyo uso no se adapta totalmente a las necesidades del profesorado en su acción tutorial. No resulta cómodo de operar por parte del profesor ya que esta diseñado para concertar una cita con un grupo, pero no para ser usados como una agenda. No genera indicadores adecuados para la gestión del sistema. Otro sistema más adecuado para la docencia sería *Appointy*. Este sistema esta orientado a la educación y es usado por colegios, universidades, bibliotecas y demás instituciones educativas que necesitan concertar citas con usuarios. Es sistema abarca otras funcionalidades adicionales a la mera gestión de citas 1 a 1 (no tiene citas de grupo) e incluye la organización, personal, eventos, pagos, etc. El servicio ofrece un plan gratuito con funcionalidades muy reducidas (máximo 1 usuario y 100 citas). La información del alumnado que maneja el sistema, al igual que con *Doodle*, recaería en servidores alojados fuera del dominio de la universidad. Finalmente, en *Moodle* se dispone de un plugin citas asociado a asignaturas (ver vídeo en [10]) con funcionalidades parecidas al sistema propuesto en este artículo. Desde nuestro punto de vista, la operativa de este plugin es excesivamente compleja para el profesorado y no reporta indicadores de gestión. Además, está ligado a un sistema concreto de enseñanza virtual que no es usado en todas las universidades. Nuestro sistema es independiente de la plataforma de e-learning usada (aunque se puede integrar en cualquier plataforma con sólo poner un enlace) y es muy simple en su operación. Finalmente, en [11] se presenta un sistema de gestión y seguimiento de citas desarrollado en el año 2013 por la Universidad Politécnica de Madrid con un propósito similar al nuestro "para facilitar y potenciar la asistencia a tutorías". El sistema desarrollado, no obstante, sólo contempla el uso de citas presenciales y parece enfocado a la reserva de turnos individuales preestablecidos (15min) y a la creación de un registro o bitácora por parte del profesorado para facilitar la evaluación del alumno. Nuestro sistema es mas flexible (p.ej, el profesor puede crear tutorías de grupo, establecer la duración de las tutorías, o crear tutorías con un propósito específico visto por lo alumnos). La aplicación desarrollada en [11] tampoco parece generar indicadores de gestión que permitan la mejora de la acción tutorial.

III. PROPUESTA DE VALOR Y REQUISITOS

Uno de los elementos clave para el éxito de un servicio TIC es la correcta identificación del valor que ofrece para los usuarios. En nuestro caso, los usuarios del servicio TIC que proponemos son tanto profesores como alumnos, ambos involucrados en el servicio de tutorías (proceso de negocio al que sirve nuestro servicio TIC) como cliente y prestador del servicio respectivamente. Según los principios de buenas prácticas en gestión de servicios [12], para ofrecer valor al estudiante, se requiere que el servicio cumpla su propósito (utilidad) y que sea usable (garantías). Respecto al componente de utilidad, no nos cabe duda de que durante una sesión de tutoría (sea presencial o telemática) el alumno suele encontrar la aclaración o consejo que buscaba. Por ello, nuestro servicio TIC centrará su aportación de valor en mejorar el proceso de negocio en todo lo relacionado con la entrega del servicio. A este respecto, al alumno le gustaría tener la tutoría con un profesor concreto en un horario que le fuese conveniente, poder elegir entre un formato tutorial presencial o telemático que se adapte a su necesidad en cada ocasión, y no tener que esperar para ser atendido. Al profesor a su vez, le gustaría tener flexibilidad para modificar el horario, y conocimiento previo sobre qué alumno desea venir, a qué hora vendrá, qué formato de tutoría quiere y qué tema desea tratar.

A. Propuesta de Valor del Servicio TIC de cita previa

Podemos identificar dos usuarios del servicio TIC de concertación de citas: alumnos y profesores. Para cada uno de estos usuarios, planteamos la siguiente propuesta de valor:

- El *alumnado* **encuentra valor** en el servicio TIC cuando consigue solicitar una cita de forma simple y rápida para el profesor que desea en un horario que le conviene sin necesidad de saber de antemano el horario de tutorías o la ubicación del despacho del profesor, de intercambiar emails con el profesor, y con la tranquilidad de que va a ser atendido sin espera.
- El *profesor* **encuentra valor** en el servicio TIC cuando puede gestionar su tiempo de forma flexible, cómoda y eficiente, tiene conocimiento con suficiente antelación sobre cuándo es la cita y qué quiere tratar el alumno sin necesidad de intercambiar emails. También encuentra valor cuando tiene indicadores que le permiten gestionar y mejorar el servicio que ofrece.

Aunque no es un usuario del servicio, podemos considerar a una unidad organizacional como cliente (p.ej., departamento, Escuela o Universidad). Este cliente encuentra valor cuando tiene indicadores que le permiten conocer la disponibilidad de los profesores, hacer un seguimiento de uso de las tutorías e indicadores que le permitan conocer el estado real del servicio que ofrecen.

B. Requisitos de la solución

Basándonos en nuestra experiencia y en los valores generales que deben guiar la gestión de los servicios según itilv4 [13] (Information Technology Infrastructure Library), planteamos los siguientes **principios de diseño**:

- *Simple y práctico.* Tanto el interfaz de usuario y la operativa de los alumnos como el del profesor deben ser lo mas simples posible.
- *Integración frente a aplicación monolítica*. No deseamos incluir funcionalidades que obliguen a cambiar las prácticas o herramientas que ya utilizan los usuarios.
- *Flexibilidad operativa frente a rigidez.* Adaptarse a las necesidades y preferencias de los profesores en la medida de lo posible.

- Poner el *valor* como principal elemento del diseño

Además de los principios anteriores, nos marcamos una serie de requisitos de diseño:

- La solución no debe tener coste económico.
- El software (servidor) y los datos deben ser gestionados y estar bajo el control del departamento o profesor.
- Se debe poder acceder al servicio desde cualquier dispositivo móvil o fijo con un navegador.
- El desarrollo, despliegue y mantenimiento deben tener un coste reducido en horas de trabajo.
- Debe generar informes e indicadores de interés para el profesor que le permitan gestionar y mejorar el servicio de tutoría.
- Debe ser no intrusiva con respecto a los sistemas preexistentes usados por el profesorado (p.ej., calendario, *BlackBoard*, *Moodle*, etc.). Por ello, la solución notificará las citas por email para que el profesor y alumno elijan la integración de la cita en el sistema de agenda que deseen.

No consideramos que dotar al servicio de mecanismos de seguridad sea necesario, ya que, en el peor de los casos, nos encontraremos citas que no se cumplen, o que se expongan las citas que tiene un profesor. Pensamos que tanto la probabilidad como el impacto de tales cosas son muy bajas.

IV. Diseño de la aplicación

En esta sección explicamos el diseño del servicio TIC. Hay dos tipos de usuarios diferenciados: profesores y alumnos. A continuación, describimos las funcionalidades por cada tipo de usuario del servicio.

A. *Usuario **alumno***

Los alumnos solicitan la tutoría con un profesor. Para ello, tienen que realizar 3 simples pasos:

1) Elegir al profesor deseado dentro de una lista.
2) Seleccionar, de entre los huecos libres ofrecidos por el profesor, el que mejor le convenga.
3) Rellenar el campo *uvus* (usuario virtual de la universidad de Sevilla), la asignatura o asunto a tratar, y pulsar el botón de solicitar.

Cualquier persona puede solicitar una cita con un profesor. Por ello, el solicitante no se registrará en la aplicación (i.e., no es un usuario de la aplicación) y no tienen que hacer ningún tipo de *login*. No obstante, en cada solicitud de cita el alumno debe escribir su identificador (*uvus*) que será parte de la información de la cita. El profesor puede configurar (según sus preferencias) que se exija autenticación del solicitante o no. En caso afirmativo, se valida el *uvus* del solicitante solicitándole su DNI.

Una vez que el alumno ha solicitado la cita, recibe un email de confirmación[1] con un enlace donde puede pinchar para cancelar la cita.

B. *Usuario **profesor***

Los profesores son lo únicos usuarios que requieren registro en la aplicación. Una vez registrado, el profesor puede realizar las siguientes tareas relacionadas con su horario de tutorías.

1) Definición de horarios de tutoría durante un periodo.
2) Consulta de tutorías pendientes de realizar.
3) Borrado de hueco o cancelación de tutoría.

Además, el profesor puede realizar las siguientes tareas relacionadas con la administración del servicio.

1) Cargar listado de alumnos de una asignatura. Si desea que los alumnos de una asignatura reciban notificaciones y puedan cancelar tutorías, se debe cargar la lista de alumnos como un fichero en formato csv.
2) Seleccionar las siguientes preferencias:
    - Citas autenticadas o no.
    - Número de días a mostrar al alumno (p.ej, mostrar los huecos disponibles en los próximos 7 o 14 días).
    - Tiempo de antelación para reserva de citas (p.ej., no se permite que un alumno reserve una cita 1 hora antes de la propia cita).
3) Consulta de estadísticas por curso académico
    - Número de citas solicitadas por hora del día, día de la semana, mes del año.
    - Número de reservas por tipo (telemática o presencial).
    - Número de alumnos por etiqueta (huecos con propósito específico).
    - Número de alumnos por asignatura.
    - Número de tutorías por alumno (*uvus*).

El profesor recibe un email con cada solicitud de cita. En dicho email se indican los datos de la cita que se hayan rellenado por parte del alumno (como mínimo *uvus* y asignatura), además de fecha y hora. El email recibido también incluye un adjunto en formato .ics para que el profesor pueda importar el evento a su calendario si así lo desea.

V. Implementación del Servicio

El servicio se ha implementado usando un motor de aplicaciones y se opera en un servidor del propio departamento.

A. *WAINE: Motor de Aplicaciones*

En una investigación previa de los autores en el contexto del desarrollo de interfaces de usuario (UI) basados en modelos, se desarrolló un motor de aplicaciones denominado WAINE [4]. La arquitectura de tiempo real de WAINE se muestra en el Figura 1, donde se puede apreciar cómo el motor de aplicaciones recibe peticiones de navegadores (clientes) y genera en tiempo real el interfaz de usuario de una aplicación concreta en función de la URI recibida en la petición. Para ello, consulta un repositorio (*UI models*) con el comportamiento

[1] Sólo si el profesor ha cargado la lista de alumnos en el sistema.

e interfaces de la aplicación accedida, y una base de datos con los datos que usa la aplicación (repositorio *UI Data*).

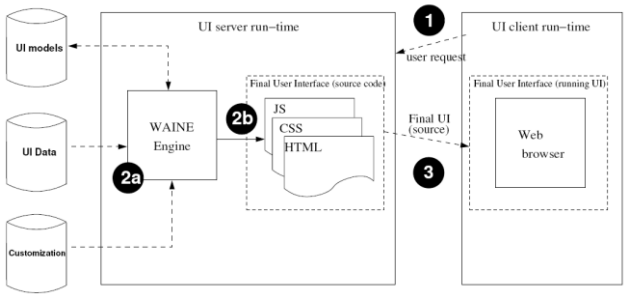


Fig. 1. Arquitectura del motor de aplicaciones

El proceso de desarrollo de aplicaciones en WAINE es sencillo. Los desarrolladores sólo tienen que especificar los modelos de una aplicación en un lenguaje XML diseñado a tal efecto, así como el esquema de la BD que usará la aplicación. Para más información se invita al lector a examinar [4].

Para nosotros, la principal ventaja de este sistema es la alta reutilización de componentes ya desarrollados en proyectos previos. WAINE incorpora varios mecanismos de reutilización que pueden acortar significativamente el tiempo de desarrollo. Otra ventaja es el bajo consumo de recursos computacionales y la facilidad de instalación (para ponerlo en ejecución sólo se requiere un servidor web + *php* + *sqlite*/*mysql*). Finalmente, WAINE lleva en operación en nuestra Escuela de Ingeniería (p.ej., aplicación de reserva de aulas) desde hace más de 10 años y ha demostrado una fiabilidad superior. Debido a estas características, y a nuestra experiencia previa, hemos usado este motor para desarrollar nuestra aplicación de citas ya que cumple los requisitos de diseño (pocos recursos, fiable, bajo coste en desarrollo). En cualquier caso, este servicio de citas de tutorías podría haberse realizado con cualquier otra tecnología.

VI. Operación de la Aplicación

Para el acceso al servicio basta escribir en un navegador la URL http://waine.us.es/apps/tutoria/. En dicho enlace, el alumno puede solicitar la tutoría con uno de los 8 profesores del Departamento que utilizan este sistema. Se invita a lector a solicitar una tutoría. En la figura 2 se muestra el interfaz que vería un alumno desde un teléfono móvil Android. El alumno visualiza todas las citas disponibles para un profesor, después sólo necesita escribir su *uvus*, el asunto y hacer *click* en el hueco que mejor le venga. También puede elegir si solicita la tutoría presencial o telemática (por defecto). La simplicidad es un valor de diseño.

Fig. 2. Ejemplo del interfaz visto por un alumno

Tras la solicitud de tutoría, el alumno recibe un email de confirmación con un enlace que le da la posibilidad de cancelar la tutoría.

La única tarea que obligatoriamente debe realizar un profesor es crear huecos horarios para que los alumnos puedan solicitar tutorías. Esto se realiza de forma sencilla a través de un formulario donde se selecciona el día y hora de inicio y final del periodo de tutoría, la duración de un turno de tutoría (en minutos) así como el número de alumnos que podrían reservar el mismo turno (p.ej. más de 1 si se desea una tutoría de grupo). En la figura 3 se puede apreciar la creación de dos huecos de tutoría individual de 15 minutos todos los martes desde el 1 hasta el 22 de junio. Si se desean crear huecos con un propósito específico (p.ej., seguimiento de un proyecto, revisión de exámenes) se puede indicar una etiqueta que aparecerá asociada a cada hueco en la vista de los alumnos. El profesor puede crear tantos periodos de tutoría como desee. De esta manera un profesor puede definir su propio horario de tutoría de manera flexible.

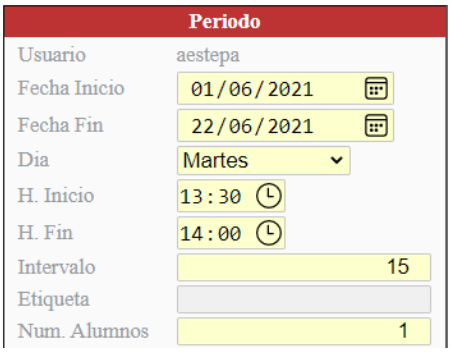


Fig. 3. Formulario usado por el profesor para crear huecos

Otra operación que el profesor puede realizar es la consulta de las citas que tiene pendientes. Esta opción es interesante si no está utilizando un sistema de calendario alternativo. En la figura 4 se puede ver un ejemplo de esta consulta de citas pendientes.

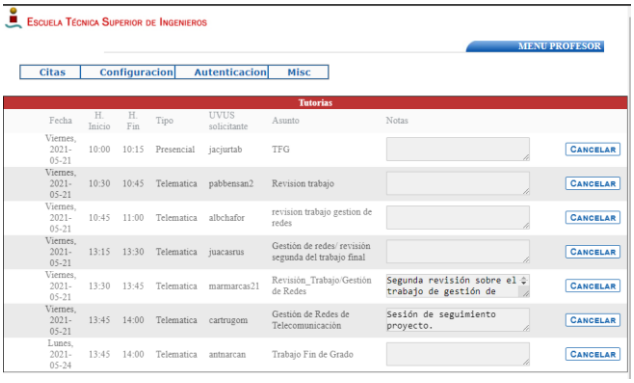
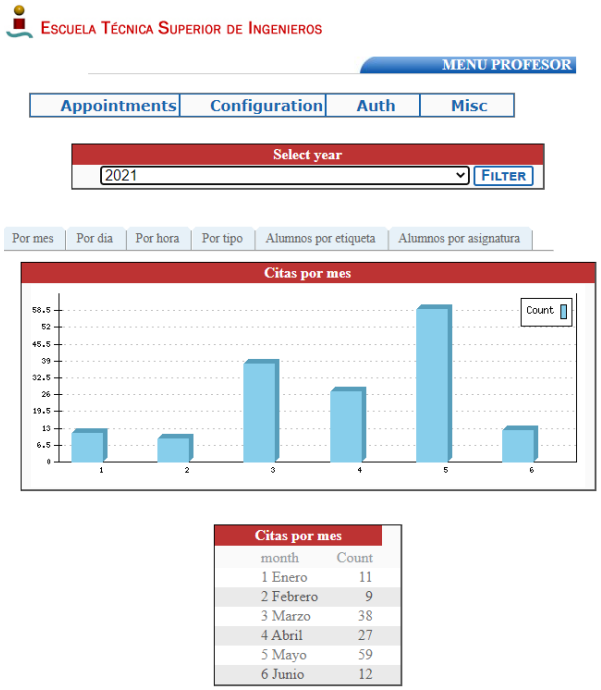


Fig. 4. Ejemplo del interfaz de citas pendientes visto por el profesor

VII. Indicadores de Gestión

El servicio de cita permite la gestión y mejora continua del servicio de tutorías a través de diversos indicadores. Los indicadores son estadísticas de uso que pueden ser consultadas a lo largo de un periodo de tiempo o durante un año académico. La figura 5 muestra un ejemplo del

interfaz de consulta del indicador de número de citas tenidas durante cada mes del año 2021.



Citas por mes

month	Count
1 Enero	11
2 Febrero	9
3 Marzo	38
4 Abril	27
5 Mayo	59
6 Junio	12

Fig. 5. Ejemplo de la interfaz de estadísticas visto por el profesor

Se han definido los siguientes indicadores.

- Número de citas por mes del año. Puede ser útil para calcular la carga de trabajo empleada a lo largo de un curso.
- Número de citas por día de la semana. Puede ser útil para averiguar cuáles son las preferencias de los alumnos (de entre los días ofrecidos por el profesor).
- Número de citas por hora. Al igual que el caso anterior, puede ser útil para averiguar qué horas del día les resultan más convenientes a la mayoría de los alumnos. Por ejemplo, en el caso del primer autor, el máximo de alumnos se produce en el horario 12-13h (84 de 155 tutorías atendidas, el 55%), mientras que 9-10 o > 16h apenas tuvo alumnos. Esto puede conducir a poner más turnos en dicho horario otros días y quitarlos de horarios menos concurridos.
- Número de citas por cada etiqueta. Una etiqueta indica un propósito específico. Esta consulta nos permitiría conocer aspectos tales como cuántos alumnos han acudido a las revisiones de exámenes o a las revisiones de los proyectos. En nuestro caso, en el periodo 2021 tuvimos 119 alumnos en tutorías sin etiquetar (ordinarias), 24 en revisión de un examen, y 13 turnos para la defensa telemática de un trabajo de una asignatura.
- Conocer individualmente las tutorías a las que ha asistido cada alumno. Esto puede ser un indicador del interés o la proactividad del alumno y podría ser un factor más a tener en cuenta para incentivar a los alumnos a usar las tutorías. A continuación, se muestra el resultado del 2021 sobre los 10 alumnos que más han asistido a tutoría con el profesor Antonio Estepa.

Tutorias por usuario (top 10)

UVUS	total
giugre	10
pabrivcar1	7
jacjurtab	6
carda1	6
roblamrod	6
antnarcan	5
albchafor	5
carvidmar	5
danneijae	4
juacasrus	4

Fig. 6. Ejemplo de resultado de consulta sobre alumnos

Además, el sistema genera un log del servicio donde quedan grabadas todas las operaciones.

VIII. OPINIONES DE USUARIOS Y PROFESORES

En la actualidad estamos aún completando un estudio estadístico sobre las opiniones de profesores y alumnos durante este curso académico. Como información preliminar, resumimos las opiniones más frecuentes en los sondeos que hemos realizado.

- Usuario Alumno. Todos los alumnos encuestados tienen una opinión muy positiva del servicio (nota media 4.4/5), creen que fomenta y facilita el uso de las tutorías. Valoran el hecho de no tener que saber el horario del profesor, la flexibilidad de elegir entre presencial o telemática y el hueco que mejor les conviene. Además, valoran el hecho de no esperar en cola en revisiones o seguimiento de proyectos. Como aspectos mejorables destacan el aspecto gráfico de la aplicación, y que a veces ninguno de los huecos libres se adapta totalmente a sus preferencias.
- Usuario Profesor. El servicio tiene algunos aspectos de configuración para adaptarse a las preferencias del profesorado (p.ej., tutorías de grupo o la exigencia de autenticación de alumnos). En general, todos se muestran muy satisfechos del sistema (4.2/5) y reconocen su influencia positiva en el servicio de tutoría. Algunos aspectos positivos comunes son la creación de un horario dinámico y adaptado a las necesidades del profesorado, así como la eliminación del email como medio para "negociar" el día y hora de una cita fuera del horario de tutoría. Algunos profesores sólo han usado el sistema para citas fuera del horario oficial de tutoría (revisiones de exámenes y seguimiento de proyectos). Otros, lo han usado para cualquier tipo de cita. Los profesores más implicados en la mejora del servicio han encontrado muy útil los indicadores de gestión. Otro efecto interesante es que los alumnos apenas han cancelado citas una vez solicitadas (menos del 1%) lo cual puede reflejar un mayor compromiso del alumno toda vez que sabe que alguien ya ha bloqueado su tiempo por él. Algunos profesores destacan la utilidad de conocer de antemano lo que cada alumno quiere tratar de cara a preparar la tutoría y hacerla mas eficiente. Por ejemplo, ha permitido saber qué alumnos acudirían a la revisión de un examen de antemano. Como aspectos mejorables destacan la necesidad de cargar las listas de alumnos (i.e., no esta conectado con el servicio

LDAP de la universidad), y la escasez de la documentación de ayuda de la aplicación.

En un año, ha habido 2 incidentes durante la operación del servicio debido a cortes del suministro eléctrico, y se han realizado 2 cambios en la aplicación debido a peticiones del profesorado, por lo que los costes en el mantenimiento y mejora del servicio han sido reducidos.

IX. Conclusiones

En este artículo hemos presentado un servicio TIC que facilita el proceso tutorial a profesores y alumnos. El servicio ofrece la concertación de cita previa a los alumnos y permite al profesorado una mejor gestión del tiempo. El sistema de citas es no intrusivo y puede integrarse con los sistemas de calendario o enseñanza virtual existentes. La opinión de los profesores y alumnos es muy positiva tras un año de uso, permitiendo un mejor aprovechamiento de la tutoría y mayor flexibilidad tanto para alumnos como para profesores. Los indicadores que genera la aplicación permiten a los profesores o gestores conocer el uso de las tutorías y mejorar la gestión del servicio tutorial.

Referencias

[1] Vick, Nicholas, et al. "The effectiveness of tutoring on developmental English grades." Community College Enterprise 21.1 (2015): 11-26.
[2] Vance, Lara Kristin. Best practices in tutoring services and the impact of required tutoring on high-risk students. Diss. Eastern Kentucky University, 2016.
[3] Kaufman, David M., and D. Bruce Holmes. "Tutoring in problem-based learning: perceptions of teachers and students." Medical education 30.5 (1996): 371-377.
[4] Zhao, Peng, et al. "Web-based medical appointment systems: A systematic review." Journal of medical Internet research 19.4 (2017): e134.
[5] https://www.softwareworld.co/top-appointment-scheduling-software/
[6] https://www.g2.com/categories/online-appointment-scheduling
[7]. Srinivas, S., & Ravindran, A. R. (2020). Designing schedule configuration of a hybrid appointment system for a two-stage outpatient clinic with multiple servers. Health care management science, 1-27.
[8] https://www.capterra.es/directory/30027/medical-scheduling/software
[9] https://www.capterra.com/appointment-scheduling-software/
[10] https://www.youtube.com/watch?v=a22Chh01zpM
[11] Cavero, Pedro Alarcón, et al. "Gestión automatizada de tutorías." REDU: Revista de docencia universitaria 12.2 (2014): 351.
[12] Van Bon, J., De Jong, A., Kolthof, A., Pieper, M., Tjassing, R., Van Der Veen, A., & Verheijen, T. (2008). Foundations of IT Service Management Based on ITIL® (Vol. 3). Van Har
[13] Agutter, Claire. ITIL® 4 Essentials: Your essential guide for the ITIL 4 Foundation exam and beyond. IT Governance Ltd, 2020.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Teoría de colas y simulación por eventos: una actividad basada en aprendizaje por proyectos

Ramón Agüero, Luis Diez
Departamento de Ingeniería de Comunicaciones. Universidad de Cantabria.
{ramon, ldiez}@tlmat.unican.es

Se describe una actividad de Aprendizaje Basado en Proyectos, que se enmarca en la asignatura de "Diseño y Operación de Redes Telemáticas", en el primer curso del Máster Universitario en Ingeniería de Telecomunicación. Los estudiantes tienen que desarrollar un simulador por eventos, y utilizarlo posteriormente para analizar el rendimiento de diferentes sistemas, cuyo comportamiento se puede estudiar en base a modelos teóricos. La actividad permite a los estudiantes adquirir un mayor conocimiento acerca de las técnicas de simulación por eventos y de teoría de colas, fundamentales en el ámbito de la ingeniería telemática. En el artículo se describen los objetivos formativos de la actividad, así como el diseño de la práctica correspondiente. La experiencia hasta el curso 2020/2021 ha sido buena, lo que se refuerza por las opiniones de los estudiantes, que valoran de manera positiva la actividad docente.

***Palabras Clave*—aprendizaje basado en proyectos (ABP), simulador por eventos, modelado, programación**

I. INTRODUCCIÓN

La simulación basada en eventos es una herramienta cada vez más usada en varios campos de la ingeniería. Por ejemplo, en el área de las redes, existen varias plataformas y herramientas ampliamente utilizadas por la comunidad para evaluar el rendimiento de protocolos, algoritmos, etc. En el ámbito docente, estas herramientas (por ejemplo ns-3[1], o OMNET++[2]) se usan con frecuencia para plantear prácticas de laboratorio, tanto en programas de grado como de máster.

Sin embargo, el uso de estas herramientas no permite habitualmente entender los detalles de la simulación por eventos, que se ocultan tras la complejidad de las implementaciones empleadas. En este artículo se describe una práctica de laboratorio planteada como actividad de Aprendizaje Basado en Proyectos (ABP), en la que los estudiantes deben diseñar y desarrollar el núcleo de un simulador por eventos, a fin de que puedan entender en profundidad el funcionamiento de esta metodología. El fin último es colocarse en una posición en la que se pueda entender mejor la operación y funcionamiento de herramientas más complejas, y cómo utilizarlas de manera adecuada para el análisis de sistemas.

Una vez implementado, el simulador se utilizará para analizar el funcionamiento de varios sistemas que pueden a su vez modelarse con cadenas de *Markov*, y cuyo rendimiento se puede por tanto estudiar mediante teoría de colas. De este modo, la actividad propuesta permite consolidar el conocimiento que se haya adquirido en el campo de dicha teoría. Además, la práctica está diseñada para incluir una serie de competencias transversales, tales como el trabajo colaborativo y la escritura de informes técnicos en lengua inglesa. La actividad se lleva a cabo en la asignatura *Diseño y Operación de Redes Telemáticas*, que se imparte en el primer curso del *Máster Universitario en Ingeniería de Telecomunicaciones* de la Universidad de Cantabria. Como se pondrá de manifiesto a continuación, la experiencia hasta ahora es positiva, ya que se consiguen cubrir los objetivos formativos, y los estudiantes tienen opiniones relativamente buenas acerca de la actividad docente.

El resto del artículo tiene la siguiente estructura. En la Sección II se mencionan algunas de las experiencias de ABP en áreas relacionadas con las redes, y se enumeran los aspectos innovadores de la actividad, cuyos objetivos de aprendizaje se identifican en la Sección III. La Sección IV describe el diseño de la práctica y detalla las modificaciones que se han ido introduciendo desde que se inició. Seguidamente, en la Sección V se comentan los resultados de las encuestas que se han llevado a cabo en los últimos tres años, las cuales muestran que, a pesar de su complejidad, los estudiantes tienen una opinión relativamente buena acerca de la práctica. Finalmente, en la Sección VI se comentan las conclusiones más relevantes que se han obtenido y se enumeran posibles modifica-

[1] https://www.nsnam.org/
[2] https://omnetpp.org/

ciones que se puedan implementar en el futuro.

II. ABP EN CURSOS DE INGENIERÍA

A lo largo de los años, la eficacia del Aprendizaje Basado en Proyectos se ha probado en varios campos [1] y especialmente en ingeniería. Una de las áreas con mayor implantación es *computer science*, tal como se describe [2] donde se resumen diversas experiencias de ABP en el área. Por ejemplo, en [3] McManus y Costello analizan la efectividad de una iniciativa similar, basada en ABP en *computer science*, tanto desde el lado de los instructores como del alumnado. Un ejemplo más concreto es el descrito por Fernandez y Williamson in [4], donde se discute cómo la metodología de ABP se usa en una asignatura de programación orientada a objetos. De forma similar, en [5] se identifican las ventajas del ABP como herramienta de aprendizaje para la implementación de servicios distribuidos.

Sin embargo, no se han encontrado trabajos previos de experiencias de ABP que cubran los elementos que se contemplan en la actividad descrita en este artículo. Como se discutirá en la Sección IV, la práctica abarca un gran número de objetivos de aprendizaje, que incluyen competencias transversales. En este sentido, mientras que la mayoría de los trabajos encontrados en la literatura se centran en el desarrollo *software*, existen muy pocos que apliquen el ABP para consolidar el aprendizaje de modelado de sistemas de comunicaciones (teoría de colas, teletráfico, etc.). Por ejemplo, Garcia y Hernandez describen en [6] los resultados obtenidos de una actividad de ABP en una asignatura de teoría de colas. Sin embargo, el estudio se centra en consejos para animar al alumnado a tomar un papel más activo a la hora de resolver ejercicios propuestos. Por el contrario, en la actividad que se describe en este artículo el enfoque es diferente, ya que el alumnado precisa asimilar el comportamiento de los sistemas para implementarlos en el simulador, de forma que posteriormente puedan comparar el comportamiento de la implementación con el análisis teórico. Este enfoque permite asimilar y profundizar los modelos teóricos subyacentes.

III. CONTEXTO Y OBJETIVOS DE LA ACTIVIDAD DE ABP

A. Contexto

Esta actividad se realiza en el primer curso del *Máster Universitario en Ingeniería de Telecomunicación*, en la asignatura llamada *Diseño y Operación de Redes Telemáticas*. En esta asignatura se tratan conceptos relacionados con el modelado analítico de varias técnicas aplicadas en redes, incluyendo algoritmos de selección de acceso, esquemas de *scheduling* y protocolos de transporte.

La asignatura se ha impartido durante 7 años, y el número medio de estudiantes es de 18. Merece la pena resaltar que todos los estudiantes de esta asignatura provienen del *Grado en Ingeniería de Tecnologías de Telecomunicación*. Este programa tiene una duración de 4 años, y se establecen 3 menciones, por lo que no todo el alumnado parte con los mismos conocimientos y competencias. En concreto, una de estas menciones es la de telemática, en la que se cubren aspectos relacionados con el modelado de redes, mientras que los alumnos provenientes de las otras opciones poseen un menor conocimiento en este área. A pesar de ello, todo el alumnado ha cursado una asignatura de introducción a la teoría de colas en el segundo curso del grado.

B. Objetivos de aprendizaje

El primer objetivo de la actividad está relacionado con la programación. En concreto los estudiantes han de implementar el simulador por eventos usando el lenguaje de programación `C`, ya que es el lenguaje que, de manera común, todos han utilizado durante el grado. Sin embargo, la adaptación de la actividad a otros lenguajes no supondría cambios en su diseño y algunos de los lenguajes alternativos simplificarían la implementación del simulador, al contar con librerías especializadas. En cualquier caso, la competencia en la programación de *software* es un aspecto clave en la ingeniería de telecomunicación y con esta actividad los estudiantes la refuerzan de manera notable.

El aspecto principal de la actividad es la metodología de análisis basada en la simulación por eventos, que es una de las piezas clave de muchas herramientas de simulación usadas en redes, así como en otros campos de la ingeniería. Dado que los estudiantes han de implementar completamente el simulador, deben abordar muchos aspectos, que al utilizar entornos de simulación existentes se dan por supuestos: (1) generación de variables aleatorias; (2) gestión dinámica de memoria; (3) generación de resultados. Con todos estos elementos, los alumnos deben realizar un análisis tipo *Montecarlo*, entendiendo los conceptos de precisión estadística, nivel de confianza. Con todo ello, la actividad abarca un número relativamente alto de conceptos fundamentales en la ingeniería.

Por otro lado, merece la pena destacar que la actividad que se describe no aborda únicamente elementos técnicos, sino que también involucra competencias transversales. En primer lugar, la práctica se realiza en grupos de 3/4 estudiantes, por lo que el trabajo colaborativo es relevante. Además, uno de las características mas distintivas del ABP es el auto-aprendizaje, ya que es muy frecuente que los estudiantes no finalicen la práctica durante las sesiones planificadas en el laboratorio. En este sentido, el uso de vídeos complementarios ha resultado ser un mecanismo eficiente para fomentar el auto-aprendizaje, como se ha mostrado en otras asignaturas [7]. Finalmente, al finalizar la práctica, los estudiantes tienen que escribir un informe técnico en el que discuten los resultados obtenidos y la estructura del simulador, lo que refuerza esta competencia, así como la lingüística, ya que el informe debe ser escrito en inglés. Por otro lado, se ha incluido recientemente el requisito de escribir el informe en LaTeXque, aunque es popular en el ámbito de la investigación, resulta desconocido para muchos estudiantes. El uso de este sistema de

edición de textos también refuerza el auto-aprendizaje, ya que se dispone de gran cantidad de información para resolver dudas.

A modo de resumen, a continuación se enumeran los objetivos formativos cubiertos por la práctica:

- Profundizar en la programación en C, que ha sido el lenguaje vehicular en el Grado en Ingeniería de Tecnologías de Telecomunicación.
- Metodología de análisis basada en la simulación por eventos, utilizada en numerosas plataformas y herramientas de simulación en el ámbito de la ingeniería de telecomunicación.
- Modelado de sistemas de telecomunicación vistos anteriormente en la titulación de grado
- Redacción de documentos científicos mediante LaTeX, incluyendo elementos matemáticos.

Asimismo, durante la realización de la práctica se potencia un número importante de competencias transversales, que que se indican a continuación

- Trabajo colaborativo al realizarse en grupos de tres personas.
- Competencia lingüística, dado que la memoria se tendrá que escribir en inglés y en algunos casos será necesario además acudir a fuentes bibliográficas en inglés.
- Autoaprendizaje, ya que la actividad docente está planteada para que se pueda desarrollar de manera completa en modalidad no-presencial aunque se contemplan sesiones de tutoría.
- Redacción de informes científicos, al evaluarse a partir de un informe generado por los grupos informe.

IV. DISEÑO DE LA ACTIVIDAD DE ABP

La práctica empieza con un seminario, donde el profesorado presenta las técnicas de programación necesarias para implementar el simulador por eventos, y que son desconocidas para una parte del alumnado: (1) gestión dinámica de memoria; (2) listas enlazadas. A partir de estos conceptos se presenta la estructura básica del simulador, incluyendo aspectos clave como son: significado de un evento, u ordenación temporal. A modo de resumen, se plantea la utilización de un bucle infinito, de forma que en cada iteración se gestiona el primer elemento de una lista de eventos, que se encuentra siempre ordenada en función del tiempo de ocurrencia en orden ascendente. Cada uno de los eventos puede dar lugar a nuevos eventos, que se introducen ordenadamente en la lista, y el procesado continua hasta que se cumple algún criterio de parada, tal como tiempo de simulación o número de eventos procesados.

A. Tarea inicial: M/M/1

La primera tarea de la práctica consiste en usar el simulador para estudiar el rendimiento de un sistema M/M/1, lo que permite validar el funcionamiento del propio simulador. Todos los estudiantes han estudiado previamente este sistema, ya que forma parte de la asignatura comunes del segundo curso del grado [8], por lo que deberían ser capaces de caracterizar su comportamiento de forma teórica.

En este sistema existen dos tipos de eventos que se han de considerar: (1) *llegada* de un nuevo paquete o petición; (2) *salida* de un paquete o petición del sistema. Al procesar un evento de *llegada*, lo primero que se debe implementar es el siguiente evento de *llegada*, usando una instancia de una variable aleatoria. Posteriormente, se debe comprobar si hay recursos libres para procesar la petición. En caso de que así sea, se debe ocupar el recurso e iniciar el procesado, por lo que se debe generar un evento (nuevamente con un instancia de una variable aleatoria) de salida del sistema en el instante en el que termina el procesado. Por otro lado, si el recurso está ocupado, la petición se guardará en una cola, a la espera de que pueda ser procesada.

Por otro lado, ante un evento de *salida*, el sistema debe comprobar si hay elementos esperando en la cola. En ese caso, se inicia un nuevo procesado, creando el evento de *salida*, mediante la realización de la variable aleatoria correspondiente.

Asimismo, los estudiantes también necesitan almacenar información durante la simulación para caracterizar el sistema, de modo que se pueda comprobar el funcionamiento simulado con el rendimiento teórico esperado. A partir de la información almacenada pueden hacer uso de cualquier herramienta matemática (Matlab, R, Octave) para procesar la información, y representar de manera gráfica el rendimiento del sistema.

B. Sistema complejo

Una vez que los alumnos han validado el correcto funcionamiento del simulador mediante el análisis del sistema M/M/1, se les pide estudiar el comportamiento de otro sistema, que es diferente para cada grupo. En los últimos años este sistema se ha tomado de problemas de cursos anteriores, lo cual tiene varios beneficios. En primer lugar, los alumnos refuerzan su conocimiento sobre el modelado de sistemas y teoría de colas, siendo este uno de los principales objetivos de la actividad. Por otro lado, para todos los sistemas propuestos [9] se cuenta con las soluciones numéricas de los mismos, de modo que los diferentes grupos pueden comprobar la validez de los resultados obtenidos mediante el simulador. En algunos casos también hay disponibles vídeos que describen de forma detallada el modelo a implementar, lo cual ha sido muy bien valorado por parte de los estudiantes [7].

Como ejemplo ilustrativo se va a presentar uno de los sistemas que ha sido asignado en el último curso y que se corresponde con el Problema 52 de [9], cuya cadena de *Markov* se muestra en la Figura 1. Se trata de un sistema de cómputo que recibe peticiones de dos tipos: α y β. El primer tipo de peticiones pueden ser procesadas de forma individual, mientras que las pertenecientes al segundo tipo tienen que procesarse junto con una de tipo α. Por lo tanto, si llega una petición de tipo β cuando el sistema está vacío (no hay ninguna petición siendo procesada), esta esperará hasta la llegada de una de tipo

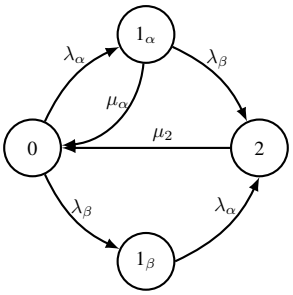


Fig. 1: Cadena de *Markov* del ejemplo de sistema complejo

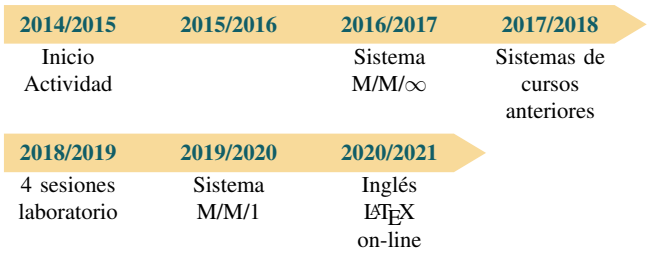


Fig. 2: Modificaciones de la práctica de laboratorio desde su inicio

α y ambas se procesarán juntas. Por otro lado, si una petición de tipo β llegara cuando se está procesando una de tipo α, el procesado se detendrá y ambas serán procesadas conjuntamente. Cuando el sistema se encuentra procesando dos peticiones conjuntas (una de cada tipo), las peticiones que lleguen no se podrán almacenar. Además, si llegara una petición de tipo α cuando otra del mismo tipo se está procesando, la nueva se descartará. En todos los casos los tiempos de procesado tienen una distribución exponencial con diferente valor medio para el procesado individual, μ_α^{-1}, y conjunto, μ_2^{-1}.

En este sistema, los estudiantes pueden analizar la probabilidad de pérdida de cada uno de los tipos de petición, así como el tiempo medio de permanencia en el sistema, diferenciado por tipo. También se podría analizar el porcentaje de peticiones de tipo α cuyo procesado se reinicia ante la llegada de una de tipo β. Como se puede observar, hay varios aspectos que pueden ser analizados con el simulador y a partir del modelo del sistema, indicado en la Figura 1, y muchos de ellos pueden ser validados con los resultados teóricos. Si los estudiantes quieren profundizar más en el modelado del sistema, tienen disponible un vídeo explicativo [10]. Cabe destacar que los enunciados de los sistemas complejos establecen un análisis mínimo que cada grupo tiene que realizar, pero se deja a su elección extender el estudio del sistema con el simulador.

C. Seguimiento de la práctica

Como se ha comentado anteriormente, para completar la práctica cada grupo tiene que redactar un informe técnico. La estructura se deja a elección de cada grupo, aunque en el guion se les indica un conjunto de resultados mínimos que deben obtener para los dos sistemas (M/M/1 y complejo), y que deberán comparar con los teóricos. Por otro lado, el informe se tiene que escribir en inglés. Aunque este aspecto podría no ser relevante, el máster en el que está enmarcada la asignatura se imparte completamente en castellano, con pocas tareas en inglés más allá de la búsqueda bibliográfica. Además, se ha incluido el uso de LaTeXcomo requisito, de modo que los estudiantes que no lo han utilizado antes deben familiarizarse con un entorno de procesado de textos diferente. A fin de simplificar la edición del informe, el profesorado proporciona una plantilla y un proyecto en *Overleaf* a cada uno de los grupos. Como se comentará en la Sección V, estos cambios han sido bien recibidos por el alumnado.

Uno de los aspectos clave de la práctica es el seguimiento continuo que se debe llevar a cabo. Debido a la complejidad del conjunto de tareas que tienen que afrontar, algunos grupos necesitan varias sesiones de tutoría, tras completar las sesiones de laboratorio programadas, en las que se resuelven dudas o se solucionan problemas concretos relacionados con la programación. Esto supone una carga de trabajo adicional para el profesorado, lo que precisa una correcta planificación, pero que en base a las experiencias previas es difícil de evitar sin modificar el enfoque de la práctica.

D. Evolución del diseño de la práctica

Desde que se inició la tarea se han ido realizando modificaciones en su diseño, basadas en la experiencia y la percepción y opinión del alumnado. A modo de resumen la Figura 2 muestra los cambios que se han ido introduciendo durante los 7 años en los que se ha llevado a cabo la actividad docente. Durante los primeros 2 años no se pedía evaluar un sistema inicial sencillo, sino que el simulador se utilizaba para analizar técnicas acceso a un medio compartido (CSMA, CSMA/CD, etc). En la tercera edición, antes de abordar un sistema de mayor complejidad, se pidió analizar un sistema M/M/∞, para que los grupos se familiarizaran con la generación de variables aleatorias exponenciales que necesitarían posteriormente. Sin embargo, el análisis del rendimiento de este sistema, que asume recursos infinitos, no permite muchas opciones, y el procesado de los eventos resulta muy sencillo. Esto daba lugar a un salto de complejidad notable con el sistema avanzado.

En el cuarto año se plantearon como sistemas avanzados ejercicios de cursos anteriores, lo que ha permitido poder profundizar en los aspectos de modelado. En este sentido, se había comprobado que los cambios introducidos hasta el cuarto año aumentaron la complejidad de la práctica, por lo que se decidió añadir una sesión adicional de laboratorio para compensar este aspecto, al menos parcialmente. Por otro lado, tras observar las limitaciones del sistema M/M/∞ como implementación inicial, se decidió cambiarlo por un sistema M/M/1, que además de ser más completo en tipos y lógica de eventos, también era conocido por todos los estudiantes. Finalmente, en el último curso (primer semestre del año académico 2020/2021) se

Tabla I: Resultados académicos de la actividad docente desde que se puso en marcha

Año	# Estudiantes	# Grupos	Nota		
			min	*avg*	*max*
2014/2015	9	3	7	8.2	9.5
2015/2016	21	6	5.5	8.1	10
2016/2017	19	6	7.5	8.6	10
2017/2018	16	5	6	8.3	10
2018/2019	22	7	6.5	8.4	10
2019/2020	23	8	7	8.4	10
2020/2021	17	5	6.5	8.5	10

ha puesto más énfasis en el informe que debe ser entregado al finalizar la práctica, que tiene que estar escrito en lengua inglesa y utilizando LaTeX. Además, debido a las restricciones impuestas por la situación de pandemia, se adaptaron las sesiones de laboratorio a modalidad on-line, lo que ha tenido una buena acogida tanto por parte del alumnado como de los profesores.

V. OPINIÓN DEL ALUMNADO

La actividad se ha llevado a cabo durante 7 ediciones de la misma asignatura, cuya información más relevante, en cuanto a calificaciones y alumnado, se resume en la Tabla I. Tras las 3 últimas ediciones se ha realizado una encuesta para recoger la opción de los estudiantes sobre la práctica, a fin de conocer su punto de vista e identificar mejoras potenciales para adaptar el diseño o ejecución de la práctica. El número total de estudiantes durante los últimos 3 años ha sido 62 y casi el 90% ha completado la encuesta de manera anónima, y tras finalizar la actividad docente.

En la Tabla II se enumeran las cuestiones que se han incluido en la encuesta. Como se puede ver el conjunto de cuestiones se han divido en 4 grupos diferentes. El primero busca conocer la opinión general del alumnado acerca de la práctica. El segundo grupo se centra en la metodología y la planificación de las sesiones de laboratorio. A continuación se ha preguntado acerca del informe técnico que deben realizar al completar la práctica. Finalmente, el último grupo de cuestiones trata de conocer la percepción de los estudiantes relativa al esfuerzo, relacionando las respuestas con su conocimiento previos en los aspectos que forman la base de la práctica: modelado de sistemas y programación. En todas las preguntas las respuestas se han numerado entre 1 (muy poco/completamente en desacuerdo) y 5 (mucho/completamente de acuerdo). A continuación se analizarán los resultados obtenidos por medio de las encuestas. Este análisis se llevará a cabo agregando todas las respuestas obtenidas a cada una de las cuestiones. En este sentido no se ha estudiado la evolución temporal de las respuestas a las preguntas, dado el limitado número de cursos (3 años) en los que se han realizado encuestas, este aspecto se abordará en el futuro.

La Figura 3 muestra la distribución de las respuestas relativas a la opinión general. Se puede observar que se tiene una opinión positiva, ya que los resultados a las preguntas Q1 y Q2 son muy altos, 4.25 y 4.26 respectivamente. También se puede ver que el porcentaje de respuestas

Tabla II: Cuestiones incluidas en la encuesta

#	Cuestión
	Opinión general
Q1	Los objetivos de la práctica están claros.
Q2	Considero que se han cumplido los objetivos formativos de la práctica.
Q3	Considero que el esfuerzo invertido en el desarrollo de la práctica está adecuadamente recompensado.
Q4	La práctica me ha gustado
	Planificación y metodología
Q5	Considero que el seminario de formación es suficiente para afrontar el desarrollo de la práctica.
Q6	Considero que las horas programadas de Prácticas en Laboratorio (4 sesiones) son adecuadas para encauzar el desarrollo de la práctica.
Q7	La carga de trabajo se ha repartido de manera ecuánime entre los/las integrantes del grupo.
Q8	El profesorado de la asignatura ha facilitado el desarrollo de la práctica, resolviendo las dudas que han ido surgiendo, y orientando el desarrollo adecuadamente.
	Informe técnico
Q9	Considero que la elaboración de la memoria final es adecuada en el planteamiento general de la práctica como complemento a la evaluación de la misma.
Q10	Considero que la redacción del informe técnico de la práctica en inglés ha supuesto un esfuerzo adicional respecto al uso del castellano.
Q11	Cuál es tu opinión respecto a la utilización de la plataforma *Overleaf* y de LaTeXpara la redacción del informe técnico de la práctica. [Como plataforma colaborativa para facilitar la redacción entre todos los miembros del grupo]
Q12	Cuál es tu opinión respecto a la utilización de la plataforma *Overleaf* y de LaTeXpara la redacción del informe técnico de la práctica. [Desde el punto de vista del uso de Latex para redactar el informe técnico]
	Conocimiento previo y esfuerzo
Q13	Mis conocimientos previos a la realización de la práctica han sido suficientes para afrontar su desarrollo. [Modelado, Teoría de Colas]
Q14	Mis conocimientos previos a la realización de la práctica han sido suficientes para afrontar su desarrollo. [Programación C]
Q15	El trabajo que he dedicado a la práctica fuera del horario de Prácticas de Laboratorio ha sido, en relación a dichas horas.

negativas (1 y 2) es muy bajo (2% y 0%), mientras que el de respuestas positivas (4 y 5) es casi el 90% en ambas. A continuación la cuestión Q3 busca conocer si se percibe que el esfuerzo está recompensado. El resultado medio a esta cuestión es 3.78%, siendo el 11% de las respuestas negativas, por un 60% positivas. Finalmente, la última cuestión del primer grupo, Q4, evidencia que la opinión general de la práctica es buena, con un resultado medio de 3.84. Como se puede ver en la Figura 3 solo el 13% de las respuestas han sido negativas, mientras en 73% son positivas.

Las respuestas del segundo grupo, centrado en la metodología, se muestran en la Figura 4. La primera cuestión, Q5, se refiere al seminario inicial sobre la

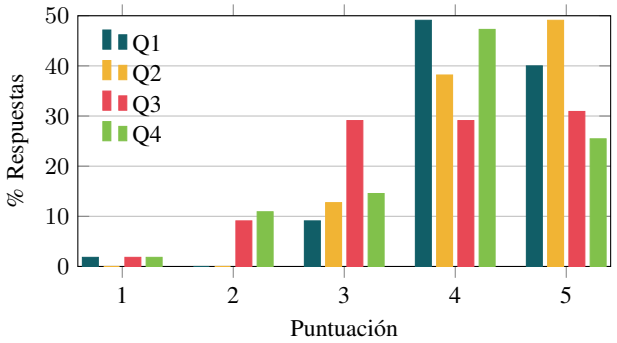


Fig. 3: Opinión general

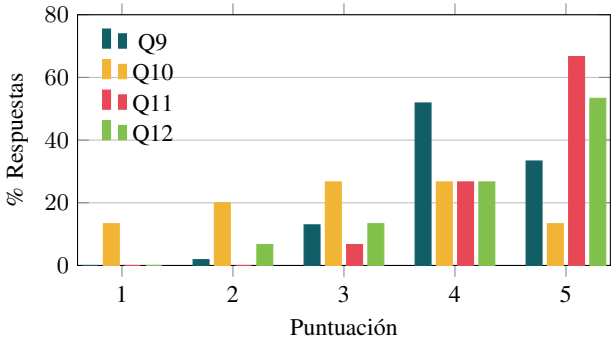


Fig. 5: Informe técnico

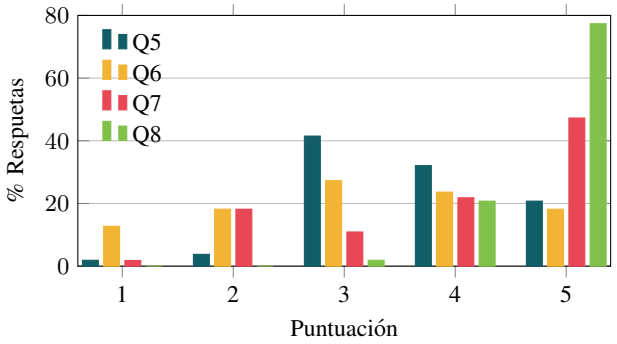


Fig. 4: Planificación y metodología

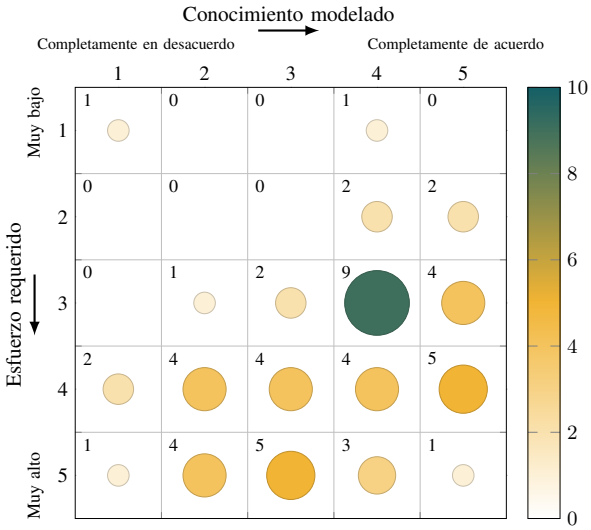


Fig. 6: Conocimiento de modelado Vs. esfuerzo

implementación del simulador. Como se puede ver, la percepción es positiva, con un resultado medio de 3.66 y siendo el número de respuestas negativas del 6%. Sin embargo, las respuestas positivas (4 y 5) solo llegan al 53%, lo que lleva a pensar (junto con comentarios en respuestas de texto libre no incluidas) que los estudiantes agradecerían un tratamiento más en profundidad de ciertos aspectos. Aunque la extensión del seminario (2 horas) no se puede extender por limitación de tiempo en la asignatura, se generará material adicional que complemente lo explicado en el seminario. A continuación, la cuestión Q6, que pregunta sobre el número de sesiones de laboratorio, obtiene una puntuación más baja (3.16) y muestra que casi un tercio de los estudiantes considera que no son suficientes.

Respecto al reparto de la tarea entre los integrantes de los grupos (Q7) la puntuación obtenida es relativamente alta (3.95), aunque el 20% de las respuestas son negativas. Para solventar esta situación, en el futuro se pedirá incluir en el informe la contribución de cada miembro del grupo, para resaltar la importancia del trabajo en equipo. La última pregunta de este grupo, Q8, es relativa al apoyo y seguimiento por parte del profesorado. Como se puede ver, los estudiantes tienen muy buena percepción, con un resultado medio de 4.75, sin que haya ninguna respuesta negativa de las más de 60. La respuestas a las cuestiones Q9-Q12 ha sido también positiva, como se puede ver en la Figura 5. En concreto, la escritura del informe se ve adecuada, aunque la imposición de hacerlo en lengua inglesa aumenta la complejidad. Por otro lado, el uso de LaTeXy *Overleaf* ha tenido una buena acogida.

Para representar las respuestas relativas al último grupo de cuestiones se ha usado una representación diferente. En la Figura 6 se muestra la distribución de las respuestas asociando el esfuerzo percibido con el conocimiento previo referido a modelado de sistemas. En cada casilla se indica el número de respuestas, que recibieron las puntuaciones correspondientes de ambos aspectos, tanto con un círculo de diferente tamaño como de forma numérica. Como se puede apreciar, el mayor número de respuestas corresponde a la combinación: esfuerzo requerido 3; conocimiento de modelado 4. Por otro lado, los estudiantes perciben que han de realizar un gran esfuerzo para completar la práctica, con un resultado de 3.71 y un 60% de las respuestas indicando que el esfuerzo es alto o muy alto. Finalmente, no se aprecia una clara relación entre el esfuerzo requerido y el conocimiento previo de modelado de sistemas.

En la Figura 7 se usa la misma representación para analizar la relación entre esfuerzo requerido y conocimientos de programación. En este caso se puede ver la relación entre ambos aspectos, de modo que aquellas respuestas que indican una menor destreza en programación también perciben que necesitan realizar un mayor esfuerzo. Se puede observar que los estudiantes consideran que iniciaron la actividad con conocimientos medios o altos de

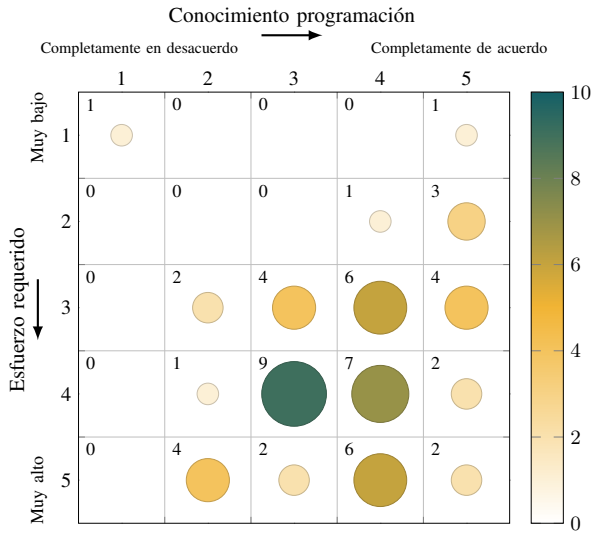


Fig. 7: Conocimiento de programación Vs. esfuerzo

programación (3.64) y que el 58% considera que sus conocimientos en este sentido eran adecuados al iniciar la práctica, mientras que el 15% creen que eran insuficientes. Junto con las respuestas relativas al seminario, estos resultados llevan a considerar la generación de material adicional relacionado con la programación del simulador, tales como un documento de ayuda o vídeos.

VI. CONCLUSIONES

Este artículo describe una actividad de ABPs que combina la implementación de un simulador por eventos con su utilización para el análisis de sistemas de comunicaciones, cuyo rendimiento se puede analizar mediante teoría de colas. La actividad se realiza como una práctica de laboratorio perteneciente a la asignatura *Diseño y Operación de Redes Telemáticas*, perteneciente al primer curso del Máster Universitario en Ingeniería de Telecomunicación.

Los objetivos de la actividad descrita cubren tanto aspectos técnicos como transversales, tales como: trabajo colaborativo, escritura de informes técnicos y autoaprendizaje. Se ha descrito el diseño de la actividad y se han enumerado los cambios que ha sufrido a lo largo de los 7 años en los que se ha llevado a cabo. Entre otros aspectos, los sistemas que se proponen para evaluar con el simulador han incrementado su complejidad, para permitir un análisis más completo. Además, la escritura del informe técnico ha ido ganado importancia de forma paulatina.

A partir de los resultados de encuestas realizadas a más de 60 estudiantes se ha visto que, a pesar de la dificultad de la actividad, la percepción es positiva. En este sentido, cabe destacar que casi la unanimidad de las respuestas indican que, tras completar la actividad, los estudiantes consideran que sus objetivos formativos se han cumplido. Por otro lado, el resultado de la encuesta ha permitido identificar extensiones y mejoras que se irán incorporando en próximos cursos.

La actividad puede ser fácilmente adaptada a otras asignaturas o áreas de la ingeniería. En base a la experiencia adquirida, los aspectos más importantes a considerar serían:

- Los estudiantes deben percibir que el esfuerzo se ve recompensado por medio de (1) las calificaciones y (2) la implicación del profesorado durante la actividad.
- Usar elementos vistos en cursos anteriores se percibe de manera muy positiva, y ayuda a consolidar conocimientos.
- El uso de recursos adicionales, tales como videos, ha sido muy bien recibido por los estudiantes y fomenta el auto-aprendizaje.

A fin de extender los objetivos de la actividad, se están estudiando diferentes extensiones: (1) uso de artículos científicos para plantear los sistemas a analizar; (2) incluir presentaciones orales para presentar los resultados de los análisis. La primera permitiría aplicar de forma directa el diseño de la actividad a otras asignaturas que abordaran otro tipo de sistemas, mientras que la segundo añadiría otro competencia transversal a la actividad.

AGRADECIMIENTOS

Los autores agradecen la financiación de Gobierno de España (Ministerio de Economía y Competitividad, Fondo Europeo de Desarrollo Regional, MINECO-FEDER) por medio del proyecto *FIERCE: Future Internet Enabled Resilient smart CitiEs* (RTI2018-093475-AI00), y de la Universidad de Cantabria mediante el Programa de Innovación Educativa.

REFERENCIAS

[1] G. E. Veselov, A. P. Pljonkin, and A. Y. Fedotova, "Project-based learning as an effective method in education," in *Proceedings of the 2019 International Conference on Modern Educational Technology*, ser. ICMET 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 54–57.

[2] S. C. dos Santos, P. B. S. Reis, J. F. S. Reis, and F. Tavares, "Two decades of pbl in teaching computing: A systematic mapping study," *IEEE Transactions on Education*, pp. 1–12, 2020.

[3] J. W. McManus and P. J. Costello, "Project based learning in computer science: A student and research advisor's perspective," *J. Comput. Sci. Coll.*, vol. 34, no. 3, p. 38–46, Jan. 2019.

[4] E. Fernandez and D. M. Williamson, "Using project-based learning to teach object oriented application development," in *Proceedings of the 4th Conference on Information Technology Curriculum*, ser. CITC4 '03. New York, NY, USA: Association for Computing Machinery, 2003, p. 37–40.

[5] S. Machado, R. Meseguer, A. Oller, M. Reyes, D. Rincon, and J. Yufera, "On the impact of pbl-based teaching techniques in an optional course on distributed applications," in *International Conference on Engineering and Computer Education*, Nov 2005, pp. 1–6.

[6] J. Garcia and A. Hernandez, "Active methodologies in a queueing systems course for telecommunication engineering studies," *IEEE Transactions on Education*, vol. 53, no. 3, pp. 405–412, 2010.

[7] R. Aguero and L. Diez, "An on-line project based learning assignment: programming an event-driven simluator to analyze queuing-based systems," in *Proceedings of the SIGCOMM Education Workshop*. Association for Computing Machinery, 2020. [Online]. Available: http://gaia.cs.umass.edu/sigcomm_education_workshop_2020/papers/sigcommedu20-final19.pdf

[8] R. Agüero. Tema 3. teletráfico. dimensionado de sistemas. redes de comunicaciones. [Online]. Available: https://ocw.unican.es/pluginfile.php/301/course/section/239/tema_03.pdf

[9] ——. Ejercicios tema 3. teletráfico. dimensionado de sistemas. redes de comunicaciones. [Online]. Available: https://ocw.unican.es/pluginfile.php/301/course/section/240/ejercicios_tema3.pdf

[10] ——. Problema 52 del tema 3 de redes de comunicaciones. [Online]. Available: https://www.youtube.com/watch?v=5d7_oGJ-wbU


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Creating Digital Awareness

Rafael Vidal Ferré, Jesús Alcober Segura, Cristina Cervelló-Pastor,
Mª Teresa Fernández Mateos, Eduard Garcia-Villegas, José M. Yúfera Gómez
Network Engineering Department
Universitat Politècnica de Catalunya (UPC)
Esteve Terradas, 7, 08860, Castelldefels.
rafael.vidal@entel.upc.edu, alcober@entel.upc.edu, cristina@entel.upc.edu,
m.teresa.fernandez@estudiantat.upc.edu, eduardg@entel.upc.edu, yufera@entel.upc.edu

As a result of the world's population being much more digitally connected, it is increasingly relevant to acquire an adequate level of digital awareness. For this reason, the objective of this work is to introduce the competence of digital awareness in higher education programs. Specifically, this work is focused on the deployment of such competence within the Degree in Network Engineering of the Castelldefels School of Telecommunications and Aerospace Engineering of the Universitat Politècnica de Catalunya. To do this, we have planned a project structured in four phases in which the tasks to be carried out by the different participants have been defined in detail. Thus, the roles involved are: project coordinator, academic content designers, support companies for this implementation, students, as well as the Degree Coordinator and the Deputy Director of Master's Studies and Head of Quality. This is the first initiative we plan to extend to other grades or levels of study in the future.

***Keywords*—Digital awareness, Digital Identity, Digital Citizenship, Degree Competences.**

I. Introduction

Our lives have an increasingly accentuated digital component, with a very high technological dependence that has recently been increased by the COVID-19 pandemic. Thus, the telematics[1] channel is becoming more and more relevant to social life, work, leisure activities, virtual shopping, or administrative procedures. In this digital scenario, we do not always act appropriately due to the lack of knowledge or skills in some cases and to unconsciousness in others.

This digital society requires that the general population, not only Information and communications technologies (ICT) professionals, have digital skills. Spain's Digital Agenda 2025 [1] establishes digital capabilities as one of the ten priority axes, and the National Digital Competence Plan [2] seeks to respond to this challenge. This situation is what motivates the present proposal, which arises from the project "Citizen Digital Awareness" (brand "Click & Safe", https://clickandsafe.upc.edu/). This initiative aims to position the Degree in Network Engineering (Grado en Ingeniería Telemática, in Spanish), the Castelldefels School of Telecommunications and Aerospace Engineering (EETAC) and the UPC, as academic references in this field, carrying out teaching, research, innovation and dissemination of this subject. It should be noted that this proposal has the support of the UPC management board, as well as the dean of the EETAC and the Department of Network Engineering.

In this context, our proposal aims to introduce digital awareness as a transversal competence in the Degree in Network Engineering [3]. This competence seeks to train future engineers to gain awareness of the implications of using ICT. For example, in terms of privacy, students should be aware of the consequences of using one solution or another to solve problems that arise during their professional life and apply this awareness in their processes, thus becoming social referents of digital awareness. Furthermore, by creating a transversal competence, we claim to implement it in other degrees of our university, whether or not they are related to ICT.

Currently, this competence is neither part of the transversal competences recognized by the UPC [4] nor part of the specific set stemming from the degree [3]. If we look at the rest of the Spanish Polytechnic universities, the situation is not very different. To the best of our knowledge, only the UPM has a transversal competence related to ICTs, but it is focused on acquiring technical knowledge and skills and following good practices [5]. This approach to what is known as digital competence (or skill) is predominant in the academic world [6]. We

[1]Note that the term "telematics" refers to the broader concept stemming from the integration of computer science and communications, as used in France and Spain, and is not limited to the communications of a vehicular system.

have also found it in the curriculum of our future students, which is defined in [8] and [9] for compulsory secondary studies (in Spanish Enseñanza Secundaria Obligatoria, ESO) and high school, respectively. However, if we look at the European Digital Competence Framework for citizens (DigComp) [2], used as a reference in the previously mentioned National Digital Competence Plan, we find that it introduces some competences dimensions aligned with our proposal. For example, related to the environmental impact of digital technologies and their use or digital technologies for social well-being and social inclusion.

Therefore, our contribution is the definition of this new competence, as a result of the UPC's "Citizen Digital Awareness" initiative.

Its introduction into the Network Engineering curriculum is not a coincidence. Several contents associated with this competence, which are already being taught in different subjects, have been detected. The aim is to give a structure to these contents, which naturally allows a global view of the competence and work on it systematically. To this end, a series of teaching materials will be developed, which will be integrated into the subjects whose concepts related to the competence are already being worked on. In addition, new possible contents related to this competence will be identified, determining the subjects in which they could be taught.

The realization of this initiative will open new lines of future work (see Section IV). From the EETAC's point of view, once the materials developed during the first-generation of the degree studies have been applied, the possibility of introducing the perspective of digital awareness in the final degree projects is considered. In UPC's terms, the material is proposed to be used in other degrees inside the ICT area. In addition, we intend to give an international dimension to this project. In this sense, work is being done towards the participation in European projects aligned with this topic, with the aim to introduce digital awareness competence at all levels of education, whether compulsory or not.

The remainder of the document is structured as follows. In Section II, we provide the main definition and focus of the digital awareness term. Section III comprises background concepts on digital awareness, classifying them on different types. In Section IV we detail the proposal of our digital-awareness project. Finally, in Section V, we give a short conclusion of this project.

II. DIGITAL AWARENESS

Digital awareness relates to people's individual awareness of the opportunities and risks associated with the ICT and the sensible ways of using them. This term applies to different aspects: from being aware of one's online identity, digital footprint and data protection rights to healthy habits in the use of technology [10].

Since the 1990s, it was clear that information technologies would trigger an industrial revolution now called Industry 4.0, transforming both the economy and the society [11]. It is then when the concept of digital inclusion appeared to ensure that all individuals and disadvantaged groups have access to, and skills to use ICTs, and are therefore able to participate in and benefit from today's growing knowledge and information society [12]. From that moment on, solutions were formulated in different fields such as public policy, technology design, finance, and management, which would allow all connected citizens to benefit equitably as part of a global digital economy [13].

Digital awareness is thus understood as an evolution of digital inclusion which aims not only at empowering citizens in the use of technology, but also focuses on its proper use, minding the risks and following safe and sensible ways to use it. This term is gaining notoriety in the late years, as research on digital awareness in education has been led by the increase of internet usage by young people during the COVID-19 pandemic [14] [15]. Nonetheless, no formal definitions of the term have been given to this date. Hence, in the following, we elaborate on a definition based on the evolution and usage of the term in academic research as well as in public and private initiatives with a similar goal.

III. TYPES OF DIGITAL AWARENESS

Digital awareness connects several topics related to the safe and sensible use of technology. Those topics can be classified into the following types of digital awareness, as shown in Fig. 1. Nonetheless, the proposed taxonomy is subject to constant evolution as ICT's new challenges, risks and opportunities arise.

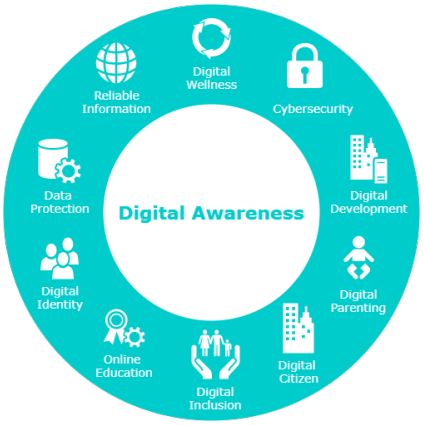


Fig. 1. Types of Digital Awareness.

A. Digital Wellness

Digital wellness (also known as digital wellbeing) is the pursuit of an intentional and healthy relationship with technology in the workplace and personal life [16].

The main issues related to the unsensible use of technology include:

- **Eye strain**. Too much screen time can force the muscles involved in eye focus and cause eye strain. Also, when looking at digital screens, the eye does not blink so frequently and this leads to dry eyes and minor eye irritations [15].

- **Insomnia**. Especially among the youngest. 77% of teenagers surveyed reported sleep problems, including night waking and difficulties falling asleep. Long gaming sessions or frequent notifications received on the phone during the night can disrupt our sleep patterns [16].
- **Diminished attention span**. Having digital devices around, especially if notifications are turned on, affects our capacity to stay focused. Also, some attention-demanding apps such as social media applications, try to build a habit on the user to check the application constantly.
- **Cyberbullying**. A group of researchers found that teenagers who spent five hours or more online each day were 71% more likely to commit suicide than those online for just one hour a day. This is related to the phenomena of cyberbullying and cyberstalking, which is especially invasive since the abuse is received via devices that accompany the victims all day [17].

Digital wellness aims to tackle these issues and promote a healthy relationship with technology. The idea is not only to learn sensible ways to use technology, but also to become aware of its use as a tool to monitor or even improve your health.

B. Cybersecurity

Cybersecurity, also referred to as information security, is the practice of ensuring the integrity, confidentiality, and availability of information. It involves a set of tools, risk management approaches, technologies, training, and best practices designed to protect networks, devices, programs and data from attacks or unauthorized access [20].

Most of the cyberattacks usually involve social engineering, which seeks to manipulate individuals to obtain confidential information or induces users to download malicious software [21].

At the individual level, cybersecurity-related digital awareness holds a set of good practices such as:

- **Setting strong passwords**. Using strong passwords combining letters, numbers and special characters is key to prevent unwanted access to our accounts and personal information. The length of the password is also key and it is advisable to use a different password in each account and pay attention to anything that may hint an unwanted external access. Changing the passwords periodically is also a recommended practice.
- **Careful use of open Wi-Fi networks**. Wi-Fi connections in public places are prone to attacks, so it is advisable not to use them, especially those not implementing security features. Check the authenticity of the pages you visit while connected to an open Wi-Fi network.
- **Do not share personal information**. Being cautious about sharing any type of personal information online is a good practice, taking special attention to social media, the enterprises and individuals that contact us.
- **Check the identity**. It is important to check the officiality of corporate accounts and veracity of individual profiles online to avoid impersonation attacks of social engineering.

It is important to raise awareness on how to identify and prevent any harmful cybersecurity threats.

C. Reliable Information

Reliable information is the information extracted from official and cross-checked sources, which contrasts with the online phenomena of fake-news and misinformation. The ability to check the veracity of online information is key to counteract the increasing viral spread of fake-news in social media [22].

The spread of fake-news can even threaten the individual safety of people, often involving health issues or economic loss. It can also affect governments degrading democracy with the phenomenon of high political polarization and sensitivity to manipulation.

D. Digital Identity

Digital identity is the set of information that identifies a person online, and it is closely related to digital footprint as it is the trace of information that a person leaves after any online activity.

It is important to understand how our actions affect our digital identity and how other people, entities, and enterprises perceive us based on our digital footprint. This can potentially affect our daily lives in unexpected ways. For instance, more than 75% of employers actively research candidates online and, what is more, around 70% have decided not to hire a candidate based on what they've found in social media; hence, online activity can even affect a person's employability [21].

Digital awareness promotes the acknowledgement of one's digital identity and digital footprint. It helps to find ways to build a solid profile, minding that the image left online will have an effect on the future.

E. Online Education

Online education regarding digital awareness relates to making use of digital resources to promote life-long learning, self-learning, and access to quality education in remote areas or during exceptional conditions.

COVID-19 pandemic crisis stimulated an unprecedented advance of the resources for online education [23]. As a consequence, education institutions foresee a hybrid education model for the near future, in which traditional and online education coexist, and adapt their programs to the digital age.

Nonetheless, there are unofficial education academias online that offer non-valid certificates and often offer misleading information and scam their clients. It is then a part of digital awareness to know how to access and take part in quality online education.

F. Digital Parenting

Digital parenting describes parental efforts and practices for comprehending, supporting, and guiding children's activities in digital environments [25].

The advisable steps to pursue good digital parenting are [23]:

- **Conversation**. Have conversations with the children on the key aspects of digital awareness such as digital footprint, cyberbullying and digital wellness.
- **Parental controls**. Use parental controls in applications to restrict inappropriate content for children and limit screen hours.
- **Becoming a role model**. Manifest digital aware habits of cybersecurity and digital wellness.
- **Celebrate**. Incentivize the good behaviours online and keep active conversation about the opportunities and advantages of the digital responsible life.

Even though academic education plays a big role in this regard, with different programmes of digital inclusion and awareness, parents should also be educated on the matter and serve as role models for their children on how to use technology safely.

G. Digital Inclusion

Digital inclusion refers to the activities necessary to ensure that all individuals and communities, including the most disadvantaged, have access to and use of Information and ICTs [27].

It must be a shared responsibility for all societies and governments to work towards digital inclusion. Both families and educational institutions must guide the youngest, the elderly and the disabled in their learning process to be digital citizens.

H. Digital Development

Digital development is the development and economic growth that comes from the use of new technologies. The ICT sector generates value and has become a new industrial revolution, Industry 4.0.

This type of digital awareness is needed so the digital transformations can be understood and supported by society. It involves acknowledging the impact of such digital transformation on all business sectors, regulating the transition, and enabling a smooth coexistence [29].

I. Data Protection

Data protection, also known as information privacy, is related to the proper handling of personal and confidential data that online entities must follow to protect it from being wrongfully shared or filtered.

It is necessary, in order to be digitally aware, to understand data protection rights and carefully read the terms and conditions agreement of every online service. Sharing one's personal data to third parties can affect online privacy and digital footprint and can even be used to manipulate or mislead users in the future.

Additionally, personal information must be used in a way that ensures the appropriate security, including protection against unlawful or unauthorised processing, access, loss, destruction or damage [30].

J. Digital Citizenship

A digital citizen is a person using ICT in order to engage in society, politics, and government. It is part of digital awareness to realize how information technologies provide a medium for easier communication with governments and a way to pursue activism to achieve a digital society that warranties safety [31].

This concept involves a full integration of digital technologies in society and has, as a political priority, the regulation of new technologies so that no legal voids are generated and new economic models do not oppress or abuse the existing ones.

This relationship between the concepts of digital citizenship and digital awareness sometimes allows the use of both terms interchangeably.

IV. Detailed proposal

As it was explained in the introductory section, the aim of the proposal focuses on the inclusion of the digital awareness competence in the bachelor's degree in Network Engineering of the EETAC.

Similarly to what we did with the rest of the competences of these degree studies, the starting point will be to define clearly and accurately the digital awareness competence, listing the different associated concepts, and mapping it in the most appropriate subjects of the degree.

The introduction of these concepts comprises two parts: (i) the creation of contents, whether theoretical or practical, and (ii) the definition of associated evaluation activities that allow assessing their level of achievement through structured grading rubrics with different criteria, levels of performance and scores.

Figure 2 shows the details of the tasks to be performed in the project, which has been structured in four phases.

The focus will be the generation of contents and assessment activities that the student can use autonomously, usually outside the classroom, but under the supervision of the corresponding faculty. This is the reason why we opt for audiovisual content, the so-called educational pills, and self-assessment tasks that can be performed on the Moodle platform.

To facilitate the incorporation of contents and evaluation activities and the subsequent use in specific subjects, the tool of the professor's "Courses Luggage" will be used. By default, we will opt for creating H5P-based interactive video allowing us to integrate audiovisual content along with its evaluation in the form of an interactive activity. All of these contents and activities will be validated and improved, if needed, according to the results of surveys conducted using Google Forms to collect feedback from the students.

There are different types of participants or roles in the project as it is shown in Fig. 3. First, there is a project coordinator. On the other hand, a working team has been defined, made up of faculty. These are academic designers of the contents. In addition, the faculty who teach the

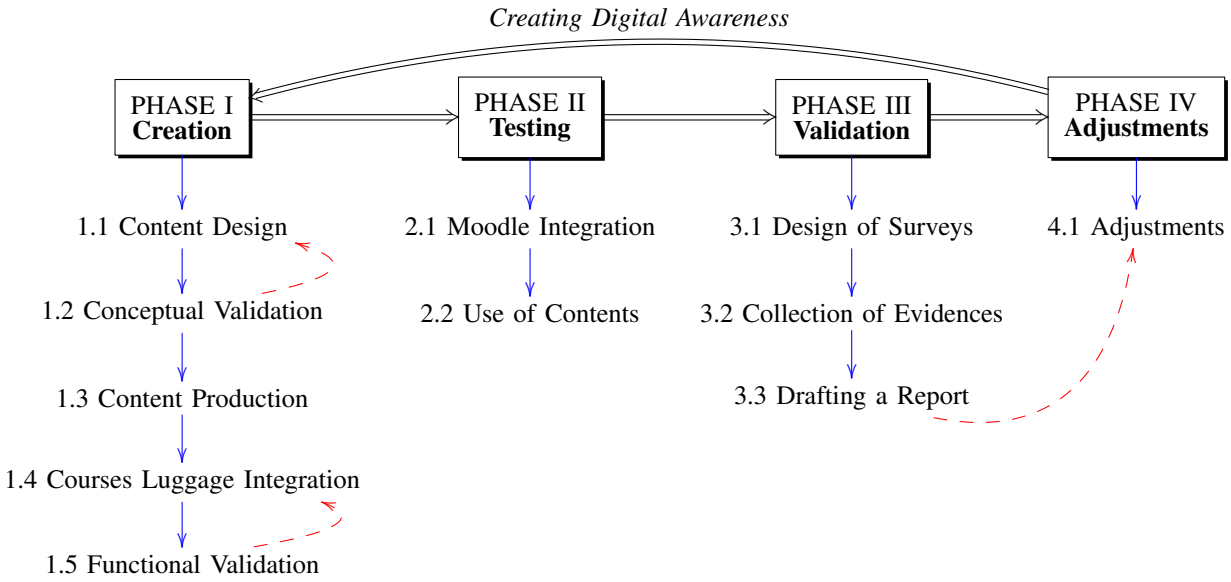


Fig. 2. Development of the project structured in four phases and several tasks. Read lines are interactions betwen them.

subjects in which the new digital awareness competence will be introduced, must also take part. Moreover, the degree coordinator and the deputy director of quality of the school also participate in the validation phase. Of course, students also participate by testing the content and answering surveys. Finally, both an external company specialized in communication strategy and UPC's production services participate in the design and production of audiovisual contents.

1) PHASE I: Creation

In the first phase, comprising several tasks, we will create different contents and the related assessment activities.

- Task 1.1. Content design: carried out by the work team in collaboration with the external company. As a result, we will obtain the video scripts and other necessary material, e.g. graphic resources, to carry out the production of the content and the definition of the associated assessment activity.
- Task 1.2. Conceptual validation: before its production, the contents and evaluation activities will be validated by the coordinator of the proposal. After production, it will also be validated by the coordinator, the academics who designed them, and by the aforementioned company. If necessary, it may involve revising task 1.1.
- Task 1.3. Content production: carried out by the audiovisual production services of the UPC and with the advice of the external company. As a result, Moodle-compatible audiovisual content will be obtained.
- Task 1.4. Incorporation of the contents into Moodle: the academics who carried out the design will be responsible for introducing the content and for creating the associated self-assessment task using the professor's "Courses Luggage" tool in the Moodle. As a result, we will get H5P-based interactive audiovisual content integrated with its evaluation.
- Task 1.5. Functional validation: to ensure that the content can be displayed correctly and that the evaluation activity is consistent with the expected answers, it will be tested at least by the academics who designed it and by the coordinator of the proposal. As a result, we will obtain the functional validation or a report of corrections to be made, which may involve revising task 1.4.

2) PHASE II: Testing

In the second phase, we will test the different contents and the assessment activities. It is divided into two tasks.

- Task 2.1. Integration in the Moodle course of the selected subject: this is a task carried out by the faculty of the subject associated with the content. As a result, the content and the associated assessment activity will be prepared to be used by students of a subject in their course at Moodle.
- Task 2.2. Use of the contents: the coordinator of the subject or the delegate will ask the students to make use of the contents that will remain available for a specific and limited period of time. As a result, we will get some positive or negative messages in the forum of the virtual campus and reports of activity and evaluation of Moddle.

3) PHASE III: Validation

In the third phase, we will test the different contents and the assessment activities. It is also divided into three tasks.

- Task 3.1. Design (or redesign) of surveys: task managed by the coordinator of the proposal, and with the validation of, at least, the coordinator of the degree and the deputy director of quality of the school. The result of this task will be a set of surveys for students and for the faculty.
- Task 3.2. Collection of evidences: surveys will be provided, through Google Forms, to the students and teachers involved in the design of the contents and the teaching of the subject. They will value the contents and, in the case of faculty, also the suitability of its evaluation.
- Task 3.3. Drafting a report: the coordinator of the proposal will write a report showing all the indicators

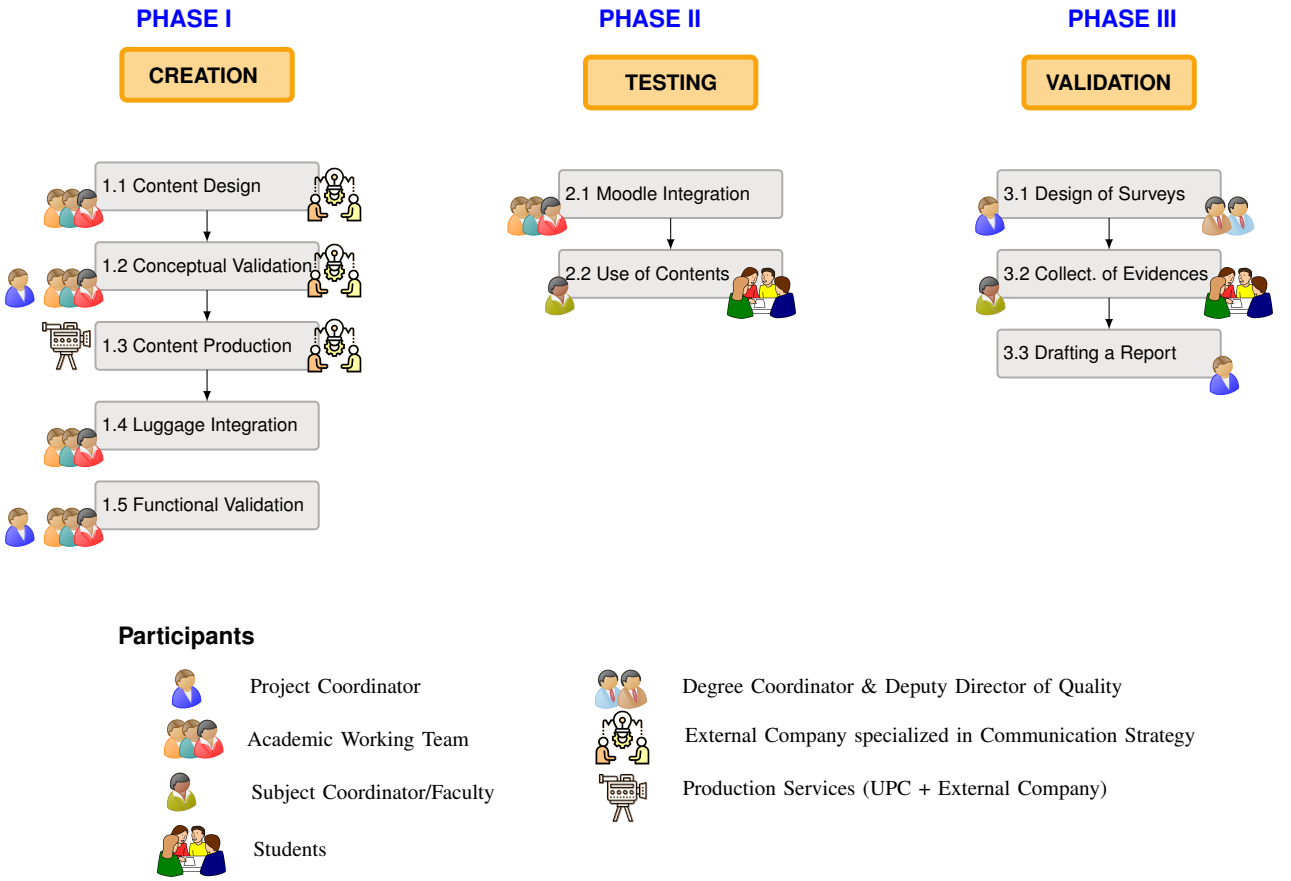


Fig. 3. Participants of the project.

collected, both in Moodle (participation in activities, evaluation results) in the form of metrics (e.g., % of participation or % of content achievement) and in the surveys. The report includes an analysis of the results, proposing, where appropriate, improvement actions and collecting feedback from the rest of the work team.

This process involves two iterations. In this way, the feedback received in the first iteration will be introduced in the second to improve the contents.

4) PHASE IV: Adjustments The fourth phase is used to adjust the contents and the assessment activities. As a result of the feedback of Task 3.3, the associated contents and assessment tasks will be adjusted, if necessary, and re-entered in Moodle.

On the other hand, in this last task, future proposals will be initiated that will take advantage of the results of this project. For example, we propose to introduce the perspective of digital awareness in the final degree projects, use it in other ICT degrees of the UPC, or participate in European projects in this field (for example with the Unite! alliance), and introduce digital awareness competence at all levels of education, whenever possible.

V. Conclusions

The objective of this work is to present an initiative related to the introduction of the digital awareness competence in the Network Engineering Degree of the Castelldefels School of Telecommunications and Aerospace Engineering (EETAC) of the UPC.

The work team is basically composed of faculty that teach in a Degree in Network Engineering. This group is based on the experience of the faculty participating in the coordination of subjects and the preparation of teaching material, apart from extensive experience in university academic management as department directors, heads of studies, department or school secretary, or section heads. The group also has experience in teaching innovation and outreach projects and in the development of the map of competencies in the EETAC degrees.

The project has been structured in four phases and several tasks planned in nine months, during which contents will be designed, developed, integrated into Moodle, and tested, evaluation included.

There are different types of participants or roles in the project, including the project coordinator, a faculty work team who will design the contents, coordinators and professors of different subjects at which this competence will be introduced, and students. It is important to highlight the roles of the Degree Coordinator and the Deputy Director of Master's Studies and Head of Quality, who will provide the project with a global perspective, both in terms of content and structure, style, and form.

In the first experiments carried out, great interest from the students has been observed, so that we expect to obtain important and remarkable results from this project.

Initially, this competence is planned to be added to undergraduate studies, but as future work, we propose to extend its application to other ICT degrees, as well as in other higher and lower study levels.

REFERENCES

[1] España Digital 2025. Portal Ministerio de Asuntos Económicos y Transformación Digital link [last accessed on Sep. 28th 2021].

[2] Plan Nacional de Competencias Digitales. Portal Ministerio de Asuntos Económicos y Transformación Digital link [last accessed on Sep. 28th 2021].

[3] Bachelor's degree in Network Engineering. EETAC-UPC, https://eetac.upc.edu/en/study/bachelors-deegrees/telematics-engineering-1 [last accessed on June 24th 2021].

[4] Transversal competences, UPC.https : / / www.upc.edu / sga / es / TitulosySET/CompetenciasTransversales [last accessed on June 24th 2021].

[5] Competencias Genéricas. Recursos de apoyo al profesorado. Uso de las TIC. https : / / innovacioneducativa.upm.es / competencias - genericas/formacion-evaluacion/uso-tic [last accessed on September 28th 2021].

[6] A. Sánchez-Caballé, M. Gisbert-Cervera, F.M. Esteve-Mon, (2020). "The digital competence of university students: a systematic literature review". https://raco.cat/index.php/Aloma/article/view/372039/465595 [last accessed on September 28th 2021

[7] Basic competences in the digital field. Identification and deployment in primary education. Generalitat de Catalunya, Departament d'Ensenyament, Nov. 2013.
http : / / educacio.gencat.cat / web / .content / home / departament / publicacions/colleccions/competencies - basiques/primaria/ambit - digital.pdf [last accessed on June 24th 2021].

[8] Core competences in the digital field. Identification and implementation in compulsory secondary education. Generalitat de Catalunya, Departament d'Ensenyament, Nov. 2015.
http : / / educacio.gencat.cat / web / .content / home / departament / publicacions/colleccions/competencies - basiques/eso/ambit - digital - angles.pdf [last accessed on June 24th 2021].

[9] High school curriculum. Generalitat de Catalunya, Departament d'Ensenyament.
http : / / xtec.gencat.cat / web / .content / alfresco / d / d / workspace / SpacesStore / 0028 / f2989dc7 - 8a2c - 4b2f - 86e8 - 4d5929f43fd7 / PUBL - curriculum_batxillerat.pdf [last accessed on June 24th 2021].

[10] International Telecommunication Union.
"Overview of ITU's History (8)" [last accessed on June 24th 2021].

[11] R. Bukht, R. Heeks, "Defining, Conceptualising and Measuring the Digital Economy," *SSRN Electronic Journal*, 2017.

[12] Communications Trust, 2020. "Digital Inclusion definition"[last accessed on June 24th 2021].

[13] Citrix Systems, Inc. "What is digital wellness?" [last accessed on June 24th 2021].

[14] I. Corradini, E. Nardelli, "Developing Digital Awareness at School: A Fundamental Step for Cybersecurity Education". In: Corradini I., Nardelli E., Ahram T. (eds) Advances in Human Factors in Cybersecurity (AHFE). *Advances in Intelligent Systems and Computing,* vol 1219, 2020.

[15] J.S. Chibbaro, L. Ricks, B. Lanier, "Digital Awareness: A Model for School Counselors," *Journal of school counseling*, 17, 2019.

[16] BetterTime, Co, 2018. "What is Digital Wellness and Why Does It Matter?" [last accessed on June 24th 2021].

[17] UAB Medicine Marketing Communications. "Eye Health in the Digital Age: Does Too Much Screen Time Hurt Your Vision?" [last accessed on June 24th 2021].

[18] C. Knight. "Screen Time and Insomnia", Feb. 1st, 2021 [last accessed on June 24th 2021].

[19] J. Byars, E. Graybill, Q. Wellons and L. Harper, "Monitoring Social Media and Technology Use to Prevent Youth Suicide and School Violence," *Contemp School Psychol*, vol. 24, pp. 318–326, 2020.

[20] J. Kaur and K.R. Ramkumar, "The recent trends in cyber security: A review," *Journal of King Saud University - Computer and Information Sciences*, pp. 1–16, 2021. Available at: link [last accessed on June 24th 2021].

[21] Social Engineering Statistics, 2021 Cyber Security Statistics. The Ultimate List Of Stats, Data & Trends, PurpleSec LLC.
https://purplesec.us/resources/cyber - security - statistics/ [last accessed on June 24th 2021].

[22] Austin Community College District, "What Makes Information Reliable?" [last accessed on June 24th 2021].

[23] L. Benedetto and M. Ingrassia, "Digital Parenting: Raising and Protecting Children in Media World," *IntechOpen chapter book*, pp. 1–21, 2020. Available at: link [last accessed on June 24th 2021].

[24] M. Fertik (April 03, 2012). "Your Future Employer Is Watching You Online. You Should Be, Too," https://hbr.org/2012/04/your-future-employer-is-watchi. Harvard Business Review. [last accessed on Sep. 29th 2021].

[25] National Digital Inclusion Alliance. Definitions, "Digital Inclusion" [last accessed on June 24th 2021].

[26] "The 7 secrets of great digital parenting". https : / / www.familyzone.com / anz / families / blog / seven - secrets - of - great - digital - parenting. The Family Online Safety Institute [last accessed on Sep 29th 2021].

[27] UNESCO, "Distance learning solutions" [last accessed on June 24th 2021].

[28] A. Seifert, PhD, S. R. Cotten, PhD, B. Xie, PhD, "A Double Burden of Exclusion? Digital and Social Exclusion of Older Adults in Times of COVID-19," *The Journals of Gerontology: Series B*, vol. 76, no. 3, pp. e99-e103, March 2021.

[29] Citrix Systems, Inc., "What is Digital Transformation?" [last accessed on June 24th 2021].

[30] GOV.UK,, "Data protection" [last accessed on June 24th 2021].

[31] M.S. Ribble, G.D. Bailey and T.W. Ross, "Digital Citizenship: Addressing Appropriate Technology Behavior," *Learning and Leading with Technology*, vol. 32, no. 1, pp. 6–11, 2004.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Sustainable Online Assessment using Interactive Multimedia Objects

Miguel García-Pineda[1], Miguel Arevalillo-Herráez[1], Esther De Ves[1], Xaro Benavent[1], Ariadna Fuertes[1], Sandra Roger[1], Máximo Cobos[2] and Diana Bri[2]

[1]Departamento de Informática,
Universidad de València.
Av. de la Universitat, s/n. 46100 Burjassot, València.

[2]Conselleria d'Educació, Cultura i Esport
Generalitat Valenciana
Av. de Campanar, 34, 46015 València

miguel.garcia-pineda@uv.es, miguel.arevalillo@uv.es, esther.deves@uv.es, xaro.benavent@uv.es, ariadna.fuertes@uv.es, sandra.roger@uv.es, máximo.cobos@uv.es, bri_dia@gva.es

In recent years, and more specifically in 2020 and 2021, multimedia elements have been the main source for students to acquire knowledge and complement their training. The aim of this work is to address the active participation of students in their evaluation process, to develop their capacity for lifelong learning. An online assessment strategy has been designed based on the use of interactive multimedia objects distributed continuously throughout the course, which ensures sustainability through the reuse of materials. The addition of interactive elements (questions, links, etc.) in a training video makes them more attractive and effective, while optimizing the teaching and learning process. In this first experience we have observed that the activity has been positively valued by the students and it has been an effective aid to learn and consolidate contents according to a high percentage of students.

Palabras Clave **online assessment, interactive video, sustainable evaluation, multimedia objects**

I. Introduction

Before the COVID-19 pandemic, 85% of youth used multimedia elements as the major source of information for knowledge acquisition [1]. During the health crisis, and still now because of uncertainty, the use of digital elements has been essential in teaching. Therefore, lecturers have had the need to design and develop new online multimedia objects that support their teaching methodologies. In this context, interactive elements (questions, clarifications, links, etc.) have gained importance, as they permit a more active participation of the student and make multimedia content more efficient than other more classical materials such as videos [2]. In addition, the effectiveness and wide acceptation of interactive multimedia elements have been shown in several previous experiences, e.g. [3].

In the present work, we describe an experience in which we address the participation of students in the evaluation process, with the intention to develop their long-term learning abilities. In particular, an online assessment strategy based on the use of interactive multimedia elements has been designed, implemented and tested, with a special emphasis on ensuring sustainability over time by make it possible to reuse materials.

Regarding the paper structure, Section II describes other related works in the same direction. Next, the suggested assessment strategy is contextualized and described in Section III. Then, the results of a student survey are presented in Section IV. Finally, section V summarizes the major findings and conclusions drawn from this work.

II. Related Work

Learning-oriented assessment is based on three basic pillars: a) a series of activities that are carried out by the students and are both meaningful and profession-oriented; b) feedback to students; and c) participation of all agents in the evaluation process through self-evaluation, hetero-evaluation and peer assessment [4]. In the context of online or blended teaching, the closer the faculty is to following a learning-oriented approach, the easier it is to plan and implement the assessment process. Along with learning-oriented assessment, other concepts such as sustainable evaluation and evaluation for empowerment shall also be considered. Sustainable evaluation is based on the application of collaborative strategies, such as self-evaluation and co-evaluation, where team members are also involved in assessment tasks. In addition, it offers students the necessary confidence to develop their learning skills throughout their lives, without increasing the workload of lecturers. In the evaluation for empowerment, the evaluation agent is no longer the lecturer. On the contrary, the student takes an active role in this process.

This implies that the students themselves must learn to evaluate and improve their performance by themselves. As a consequence, the evaluation process entails a greater complexity, which becomes even greater in the online case, when evaluation activities need to be programmed so that they appear adequately distributed throughout the course.

The design of such evaluation strategies needs to consider several factors: a) the learning outcomes; b) the knowledge and level of competence acquired; and c) the evidence, evaluation tasks, criteria, techniques, and instruments used to evaluate. In our case, our assessment will rely on videos and/or short presentations that will include interactive questions when they are played. These interactive multimedia objects will allow the system to store the answers to the questions, so that both students can use them as feedback to supervise knowledge acquisition and contribute to the continuous improvement of learning [5].

III. Sustainable Interactive Assessment

In this section, we present the interactive evaluation activity carried out in the following modules:

- Multimedia Information (MI-UV) in the Multimedia Engineering degree.
- Fundamentals of Computer Networks (FCN-UV) in the Computer Science degree.
- Databases and Information Systems (DIS-UV) in the Multimedia Engineering degree.
- Next Generation Information Systems (NGIS-UV) in the Computer Science degree.
- Image Processing (IM-UV) in the Data Science degree.

All these modules are delivered at ETSE-UV, Universitat de València (Spain). To generate the interactive multimedia objects, we have used the open-source platform H5P[1], which allows the content designer to superimpose interactions on the videos and/or presentations, e.g. images, text elaboration, links and questions. These interactive elements appear while the student is visualizing the video. H5P has been integrated into the Moodle[2] platform, allowing the creation of different types of interactive resources. To carry out the proposed activity, students were grouped in teams of *n* people, each group had to choose one of the supporting videos por the course (prepared by the lecturer), and they had to include questions, exercises and other interactive elements to produce an interactive video. Videos were divided by themes, which were related to specific concepts presented in lectures. To automate this process, lecturers used the Moodle's "group self-selection" tool, which allowed them to create as many groups as videos were available and assign students to each of these groups. For this activity, each group had to perform two different tasks:

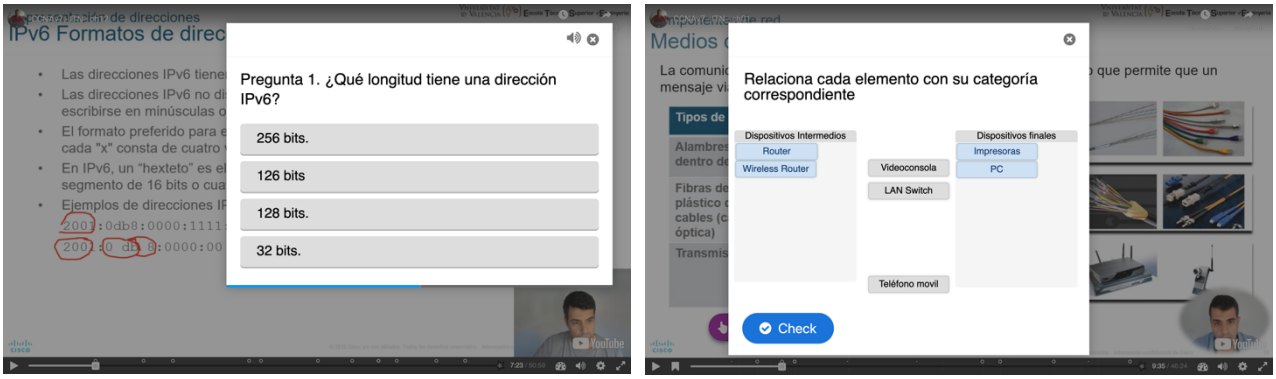
- *Proposal and creation of questions*: After visualizing the video, the group members should propose a minimum of 5 questions and embed them at the appropriate point within the video timeline.
- *Evaluation of questions from other groups*: each group, using a rubric (see Table I) provided by the lecturer, should evaluate the questions proposed by their peers from other groups. Each group will need to evaluate the questions asked by two other groups, in a peer review fashion.

This is an optional but evaluable activity. The mark for this activity is determined as a weighted average of the two activities above. The grade is a combination of the lecturer's and classmates' evaluation results, with weights 25% and 75%, respectively, and represents a percentage (between 5 and 10%) of the final grade for the module. The delivery of the activity is done through Moodle by means of a "WORKSHOP". In this workshop the interactive video in .h5p format must be submitted, that can be downloaded from the account created in the H5P platform.

Table I. Rubric for peer review process.

	Excellent (4)	Satisfactory (3)	Improvable (2)	Inadequate (1)	Score
Number of interactive elements	More than those indicated by the lecturer.	Those indicated by the lecturer.	Fewer than those indicated by the lecturer and more than 2 elements.	Equal to or less than 2 interactive elements.	
Content of the elements	The elements introduced allow to think and reflect on what is explained in the video.	The elements introduced allow a small reflection on what is explained in the video.	The elements introduced are trivial and very easy to answer.	The elements introduced are very complicated and difficult to answer with the information of the video.	
Originality	The interactive elements introduced demonstrate great originality (a wide variety of types of interactive elements). The ideas are creative and ingenious.	Interactive elements introduced demonstrate some originality. New ideas and more than 2 types of interactive elements are introduced.	Interactive elements introduced are unoriginal, at least 2 types of interactive elements are used.	The interactive elements introduced are always the same and they are not original at all.	
Use of language	No spelling mistakes or grammatical errors.	Two or less grammatical and/or spelling errors.	Between 3 and 4 grammatical and/or spelling errors.	More than 4 spelling and/or grammatical errors.	
Interest	Elements included clearly reinforce the interest in watching more videos.	Elements included reinforce the interest in watching more videos.	Elements included do not reinforce the interest in watching more videos.	The included elements detract from the interest in watching more videos.	
Total					

[1] H5P: https://h5p.org/
[2] Moodle: https://moodle.org/



a) Test question b) Drag and drop question

Fig. 1. Interactive video samples.

Once they are completed, all these interactive videos will be available in the Moodle platform (see Figure 1) as complementary material for the course. In order to encourage their visualization by classmates, a test-type exercise is included in the final exam, including some of the questions proposed by the students in the interactive videos submitted.

IV. Evaluation and Results

In order to evaluate the activities proposed in the previous section, we have used the survey shown in Table II. This questionnaire was answered by a total of 120 students, which were enrolled in one of the several subjects in which this activity was carried out. We analyze the responses globally in order to observe the degree of appropriateness and acceptance of these activities by the students.

According to the answers to question 1 (see Figure 2), 59% of the students already had prior knowledge of the assigned topic, and by performing the part of the activity that included interactive elements, they have reinforced the content of that topic or part of it. In addition, adding answers a), b) and c) we can see that 91% of the students consider that this activity has helped them to reinforce their knowledge and learn new aspects.

Figure 3 shows the responses related to question 2 of the survey. This question deals with the evaluation phase of the activity, and implicitly with the students' metacognition. In this case, 30% of the students think that they have learned new aspects about the topic covered in the video, and 61% (adding answers b) and c) of question 2) consider that the evaluation phase of the activity has helped them to consolidate their knowledge about the visualized topic.

Table II. Survey for the activity.

OBJECTIVE OF THE ACTIVITY
1. Did the activity in which you had to create questions to evaluate the assigned video help you to reinforce the knowledge you already had on that topic? a) It didn't help me at all, it was a waste of time. b) I had no idea about the topic, and by thinking and writing the questions I reinforced the content. c) I already had prior knowledge of the assigned topic, and by asking the questions I have reinforced the content. d) I have learned new aspects of the topic that I did not know.
2. Did the peer evaluation activity, in which you evaluated two videos made by your classmates on different topics, help you to encourage self-reflection (meta-cognition) in the learning of the subject? a) It has not helped me at all, it has been a waste of time. b) I had no idea about the subject of the video, and by watching the videos and answering the questions I have reinforced the content. c) I already had prior knowledge of the topic of the video, and by asking the questions I have reinforced the content. d) I have learned new aspects of the topic covered in the video that I did not know.
TIME OF DEDICATION
3. How many hours did you spend approximately to prepare the activity of thinking and writing the questions? a) 1 hour b) 2 hours c) 3 hours d) 4 hours e) More than 4 hours
SATISFACTION
4. How would you rate the proposed interactive video activity? a) Positive b) Negative

Figure 4 shows the time required by the student to prepare the activity of thinking and writing the questions of a video. It can be observed that around 67% of the students needed between 1 and 2 hours to develop the interactive evaluation activity, and only 2% required more than 4 hours to carry it out. Therefore, we can conclude that the completion time for this team-based activity did not significantly influenced the normal development of the subject or the progress of other subjects that the student may be studying at the same time.

Finally, the fourth question (see Figure 5) addresses the overall assessment of the activity. The 88% of the students considered it a positive experience, with a positive impact in learning.

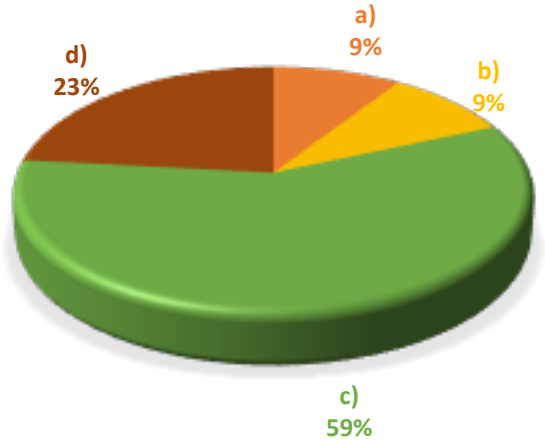


Fig. 2. Answers to question 1

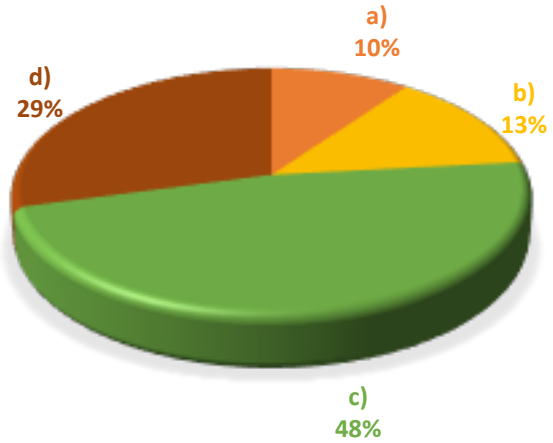


Fig. 3. Answers to question 2.

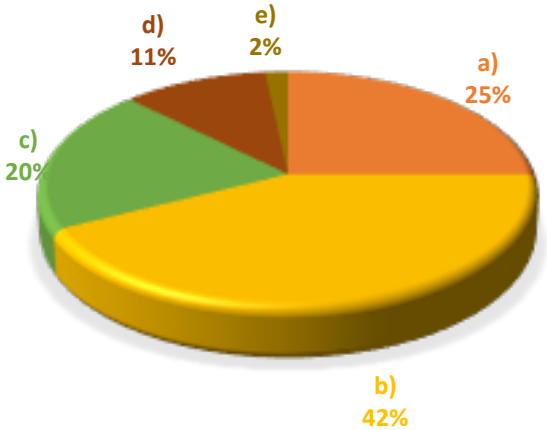


Fig. 4. Answers to question 3.

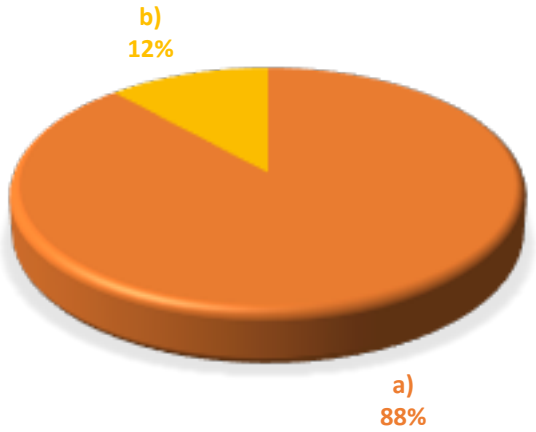


Fig. 5. Answers to question 4.

V. Conclusion

In this paper, we have presented an experience that addresses the active participation of students in the assessment process, in order to develop their learning abilities. For this purpose, an online assessment strategy based on the use of interactive multimedia objects has been designed and developed with the aim of promoting learning by explaining.

In general terms, we have observed that the activity has been positively valued by the students and has served as an aid to learn and strengthen the contents of the subjects involved according to a high percentage of students. However, and despite that these first results encourage us to further develop this type of approaches, there is still room for improvement, as it is revealed by a 12% of the students surveyed, who consider that this type of activities does not bring any improvement in the teaching-learning process.

Acknowledgements

This paper has been funded by the project SFPIE-PID20-1352382 of SFPIE at Universitat de València.

References

[1] Carolina Almeida y Pedro Almeida. Online educational videos: The teenagers' preferences. En Iberoamerican Conference on Applications and Usability of Interactive TV, pp. 65–76. Springer, 2016.

[2] Cynthia J Brame. Effective educational videos: Principles and guidelines for maximizing student learning from video content. CBE—Life Sciences Education, 15(4), 2016.

[3] Miguel Garcia-Pineda, Esther De Ves, Sandra Roger, Máximo Cobos, José M. Claver, Xaro Benavent, Miguel Arevalillo-Herráez y Juan Gutierrez-Aguado. Vídeos interactivos para mejorar el proceso enseñanza-aprendizaje en la generación YouTube. En Actas de las Jenui, vol. 5, pp. 353–356, 2020.

[4] David Carless. Learning-oriented assessment: conceptual bases and practical implications. Innovations in education and teaching international, vol. 44 (1), pp. 57–66. Routledge, 2007.

[5] Susana Olmos Miguelañez. Evaluación formativa y sumativa de estudiantes universitarios: aplicación de las tecnologías a la evaluación educativa. Tesis doctoral, Universidad de Salamanca. Ed. Universidad de Salamanca, 2008.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Codificación de vídeo basada en VMAF para escenarios DASH

Wilmer Moina-Rivera, Juan Gutiérrez-Aguado, Miguel Garcia-Pineda.
Departamento de Informática,
Universitat de València,
Av. de la Universitat, s/n. 46100 Burjassot, Valencia.
wilmoiri@alumni.uv.es, juan.gutierrez@uv.es, miguel.garcia-pineda@uv.es.

Actualmente, la mayoría de las empresas que ofrecen servicios de transmisión de vídeo a través de Internet utilizan DASH o HLS para entregar el contenido de vídeo. El uso de estas tecnologías se ha popularizado debido a la capacidad de ofrecer la mayor calidad permitida por la red a los dispositivos del usuario final, mejorando así la calidad de experiencia. Para crear una representación DASH, el vídeo de entrada de alta resolución se divide en segmentos de igual duración, y cada segmento se codifica a diferentes resoluciones. Este trabajo presenta un esquema de codificación capaz de adaptar los parámetros de codificación en cada segmento para lograr una calidad objetivo. La propuesta se ha validado con vídeos 4K y diferentes tamaños de segmento, y se ha comparado con un esquema de codificación de CRF fijo. Los resultados muestran que la solución propuesta puede reducir el tamaño del vídeo manteniendo la misma calidad.

***Palabras Clave*—Vídeo, Transmisión adaptativa, DASH, VMAF, Codificación de vídeo.**

I. INTRODUCCIÓN

Según el Informe Anual de Internet de Cisco (2018-2023) [1] casi dos tercios de la población mundial tendrán acceso a Internet en 2023. Esto significa que unos 5.300 millones de usuarios podrán conectar sus 19.000 millones de dispositivos a Internet. El vídeo, los videojuegos y el resto de elementos multimedia constituirán más del 80% de todo el tráfico, sin tener en cuenta las nuevas practicas de consumo de contenido generadas por la pandemia [2]. A todo esto se suma la aparición de nuevas plataformas de transmisión como: *Disney+, Amazon Prime Video, Twitch, Tik Tok, etc.* a las ya conocidas como *Youtube, Netflix, HBO Go*.

Hoy en día, la mayoría de estos servicios de transmisión utilizan HTTP Adaptive Streaming (HAS) como tecnología para la distribución de vídeo a través de Internet. En HAS, un vídeo se divide en trozos más pequeños llamados segmentos y cada segmento se codifica en varias resoluciones y tasas de bits. El objetivo es entregar dinámicamente el segmento de mayor calidad dadas las condiciones cambiantes al usuario final. Aunque hay muchas soluciones propietarias basadas en los principios de HAS, HTTP Live Streaming (HLS) y Dynamic Adaptive Streaming over HTTP (DASH) han logrado el éxito comercial y se han convertido en los formatos de transmisión de vídeo dominantes en Internet [3].

Pero si queremos optimizar el proceso de codificación para cada resolución y segmento, ¿qué tasa de bits o CRF (Constant Rate Factor) debemos seleccionar en cada momento? ¿Es esta selección independiente del contenido/tipo de vídeo? ¿Cuál es la mayor tasa de bits o CRF necesaria para conseguir la mejor calidad perceptible?.

Para responder a estas preguntas, presentamos un sistema capaz de codificar un vídeo a partir de un valor de métrica objetiva de calidad, como VMAF (Video Multimethod Assessment Fusion) [4]. Partiendo de un vídeo en bruto con resolución 4K, el sistema propuesto reduce la dimensión del vídeo y para cada segmento selecciona el valor de CRF que se ajusta al valor objetivo de calidad. Utilizando estos valores de CRF, el sistema codifica el vídeo a las resoluciones seleccionadas.

El sistema ha sido evaluado y comparado con una codificación basada en un valor de CRF. Los resultados muestran mejoras que oscilan entre el 1% y el 12% en cuanto al tamaño final del vídeo, manteniendo el valor de calidad indicado por el usuario[1]

Este artículo está organizado de la siguiente manera. La Sección 2 muestra varios trabajos relacionados para presentar las principales diferencias entre nuestro sistema y sus propuestas. En la Sección 3 se presenta nuestro sistema de codificación de vídeo en bucle basado en la calidad. La Sección 4 valida nuestro sistema propuesto mediante experimentos. Por último, en la Sección 5 se presentan las conclusiones y el trabajo futuro.

[1]Se puede ver una demostración de los vídeos codificados en el siguiente enlace: https://links.uv.es/jgutierr/qinloopcoding

II. TRABAJOS RELACIONADOS

A lo largo de los años, se han propuesto varios esquemas de codificación para la preparación de contenidos de vídeo bajo demanda (VoD). En 2015, Netflix propuso un enfoque [5] en el que se realizaba un análisis del vídeo para determinar la codificación óptima en función de su complejidad. Su objetivo era encontrar un equilibrio entre la tasa de bits, la calidad y la complejidad del vídeo, pero detectaron en un estudio posterior que esta solución introducía importantes artefactos de codificación. Otra propuesta de Netflix fue [6], en la que propone un sistema para controlar la tasa de bits-resolución-calidad por segmento con el fin de lograr una calidad consistente en todos los niveles, dependiendo de la complejidad del vídeo fuente. La complejidad de cada segmento de vídeo se ajustó a una tasa de bits óptima entre las diferentes regiones de calidad, pero esta solución requiere dos o tres pasadas de codificación por segmento.

Otro trabajo con una perspectiva similar es [7], en él se contemplan múltiples pasadas para utilizar las estadísticas de los pasos anteriores para alcanzar una tasa de bits objetivo en cada segmento. En este caso el uso de una relación lineal entre el CRF y la tasa de bits pretende cumplir con las restricciones de tasa de bits máxima y de calidad. En [8] se presenta un control de tasa de bits adaptable al contenido basado en redes neuronales capaz de predecir el valor óptimo del factor de velocidad para reducir las fluctuaciones de calidad en un proceso de codificación.

Debido al gasto computacional que supone alcanzar un objetivo de calidad en un proceso de transcodificación, el artículo [9] presenta una red neuronal que integra un modelo alternativo de tasa de bits. Esta propuesta tiene como objetivo estimar los parámetros de control de calidad y alcanzar un objetivo de tasa de bits en cada segmento del vídeo. También se pueden encontrar otras soluciones [10] con un enfoque diferente donde se explotan las características de la percepción visual humana para ajustar los parámetros de cuantificación (QP) durante la codificación.

Como hemos visto, los sistemas de codificación o transcodificación de contenidos de vídeo proponen sus soluciones para obtener una calidad constante con la menor tasa de bits. Sin embargo, para lograr este objetivo pueden utilizar múltiples pasadas sobre un mismo segmento de vídeo y en sus diferentes representaciones de calidad, aumentando el tiempo de procesamiento y el gasto computacional. Además, en algunos trabajos el coste de un análisis previo se sitúa entre el 41% del tiempo del segmento. Aunque se han entrenado redes neuronales en esta fase para detectar los parámetros de codificación ideales, no se ha conseguido más de un 85% de efectividad con una primera pasada. En las evaluaciones no se especifican las repercusiones que pueden aparecer si se parte de un vídeo en formato RAW en su escala original, preferiblemente 4K, ni tampoco se muestra un estudio exhaustivo de la preparación del contenido en un rango finito de resoluciones. Aspectos que sí son tenidos en cuenta en este trabajo.

III. PROPUESTA DE CODIFICACIÓN DE VÍDEO BASADO EN LA CALIDAD

Esta sección describe el proceso utilizado para la preparación del vídeo para DASH, teniendo en cuenta la segmentación para este tipo de transmisión de contenido.

La propuesta, a diferencia de otras, utiliza la calidad de vídeo, mediante la métrica VMAF, como una entrada al sistema de codificación, con el objetivo de seleccionar los parámetros de codificación adecuados para cada segmento de vídeo. Cada representación puede ser caracterizada por un valor de calidad objetivo preestablecido por el usuario, que corresponde a un valor de calidad constante en todo el vídeo. Además, estas secuencias se codifican para una gama finita de resoluciones manteniendo el mismo nivel de calidad mediante una sola pasada.

A. Algoritmo de codificación basada en calidad

En esta subsección se describe el algoritmo desarrollado y cada uno de los componentes que conforman el sistema de codificación. En ella también se explica brevemente el proceso de análisis de calidad y despliegue. El sistema de codificación se ha diseñado en dos componentes con el fin de ampliar su usabilidad y su despliegue en sistemas de procesamiento distribuido, tal y como se muestra en el diagrama de bloques (ver Fig. 1).

El primer componente, denominado **Core-Loop**, se encarga de procesar el contenido y obtener la información necesaria para codificar el vídeo a una calidad objetivo especificada por el usuario. Este componente toma como entrada el vídeo en bruto en su resolución original, 4K en nuestro caso. El vídeo en crudo, v, es procesado por el ***Módulo de Dimensionalidad***, que produce un vídeo en crudo v' a una resolución menor (por ejemplo, 240p) manteniendo la relación de aspecto del vídeo original. Después de este proceso, v' es utilizado por el ***Módulo de Segmentación*** para dividir el vídeo en segmentos de duración fija $\{v'_1, \cdots, v'_N\}$. Una vez segmentado el vídeo, en el ***Módulo de Preparación de Datos*** se codifica cada segmento v'_i a diferentes valores de CRF $\{10, 16, 22, 28, 34, 40, 46, 51\}$, para así disponer de un amplio rango de duplas $\{CRF-VMAF\}$. Todos estos segmentos codificados se pasan al ***Módulo de Calidad*** junto con el segmento original para obtener el valor de la calidad. Esto significa que para cada segmento se genera un conjunto de puntos $\{(CRF_j, Q_j)\}$. La Fig. 2(a) muestra los puntos obtenidos para un segmento junto al valor recomendado de CRF para un valor objetivo VMAF de 90 y la Fig. 2(b) todas las curvas obtenidas para todos los segmentos del vídeo Sintel[2].

El segundo componente, llamado **Per-Quality**, toma los datos generados y la calidad objetivo Q_t. Esta información se pasa al ***Módulo de Análisis de Datos*** cuyo objetivo es encontrar el valor de CRF para cada segmento con el que se obtiene la calidad objetivo. Para cada segmento es posible definir la función $Q(CRF)$ interpolando los puntos $\{(CRF_j, Q_j)\}$, como se muestra en la Fig. 2(a). A partir de esta interpolación es posible encontrar el valor

[2]https://mango.blender.org

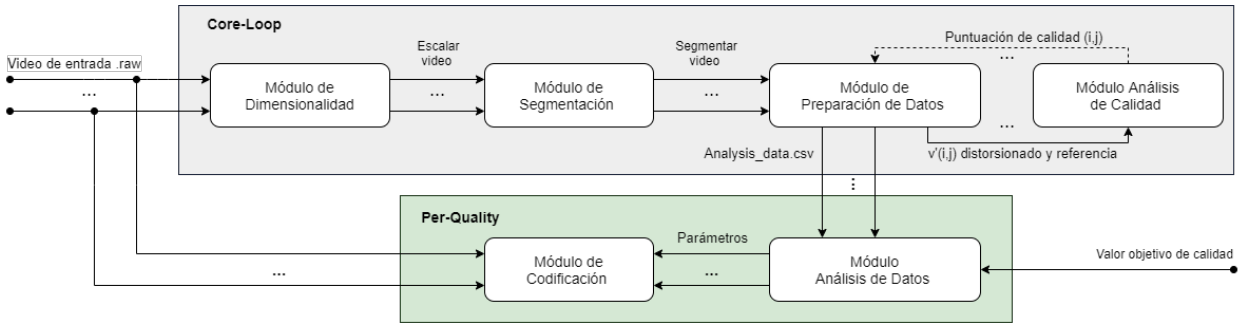
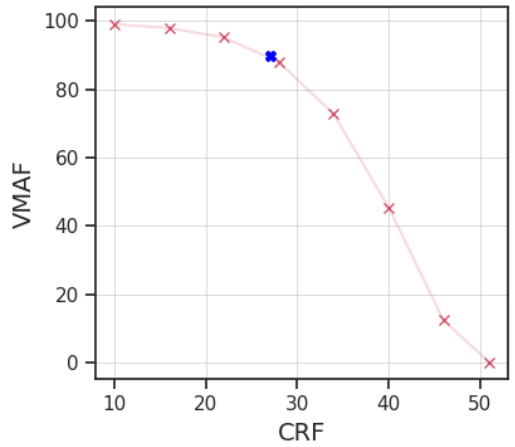
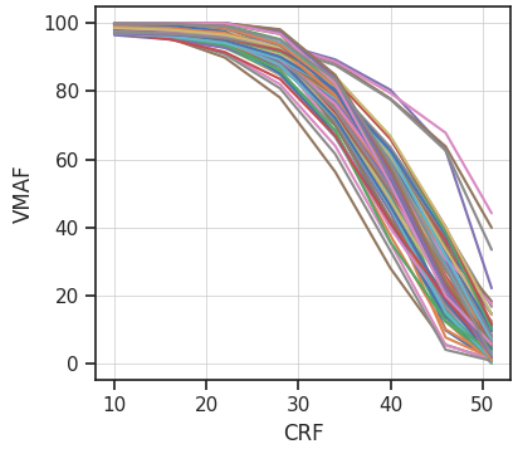


Fig. 1: Componentes y módulos para la codificación DASH basada en la calidad.



(a) Puntos muestreados y valor de CRF recomendado para un segmento.



(b) Curvas para la interpolación.

Fig. 2: Curvas utilizadas para determinar el valor del CRF ajustado a la complejidad de cada escena dado un valor objetivo de calidad.

de CRF a aplicar al segmento para conseguir la calidad obtenivo Q_t.

Una vez calculados los valores CRF, cada segmento puede ser codificado para diferentes resoluciones manteniendo la relación de aspecto del vídeo original. Para este proceso, el ***Módulo de Codificación*** utiliza como entrada el vídeo original v, el valor CRF calculado para cada segmento y el rango de resoluciones.

La solución propuesta prepara el contenido de vídeo para las transmisiones DASH bajo demanda con la calidad deseada. La calidad de las diferentes representaciones que componen la secuencia dependerá únicamente del valor de calidad objetivo definido por el usuario.

IV. EXPERIMENTOS Y EVALUACIÓN DEL RENDIMIENTO

A. *Información del Entorno de Pruebas*

En este trabajo, los vídeos se han codificado utilizando x264[3] y como métrica de calidad se ha usado VMAF[4] en su versión v2.0.0. Sin embargo, la solución propuesta puede utilizar diferentes codificadores y diversas métricas. Para evaluar y validar el método propuesto, se han realizado varias pruebas sobre dos vídeos: Tears of Steel (TOS)[5] y Sintel (STL).

Estos vídeos tienen una resolución espacial de 4K con una tasa de fotogramas de 24 fps, y están almacenados en formato YUV4Mpeg (y4m), tal y como se resume en la Tabla I. Además, para aproximar nuestra propuesta a un entorno real, la duración de los vídeos es superior a 10 minutos.

Tabla I: Características de las secuencias de vídeo utilizadas para la evaluación y la validación.

Vídeo	Resolución	fps	Fotogramas	Tamaño (GB)	Categoría
STL	4096x1744	24	21312	213	Fantasía
TOS	4096x1714	24	17620	173	Acción

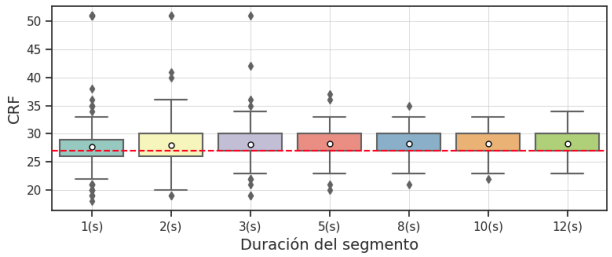
El hardware/software utilizado en todos los experimentos se especifica en la Tabla II. Se han utilizado contenedores Docker[6] con el objetivo de aprovechar la capa de abstracción y automatización en la virtualización de aplicaciones, y su capacidad de ser desplegados en

[3]ffmpeg con la biblioteca del códec libx264
[4]https://github.com/Netflix/vmaf/releases/tag/v2.0.0
[5]https://durian.blender.org
[6]https://www.docker.com/

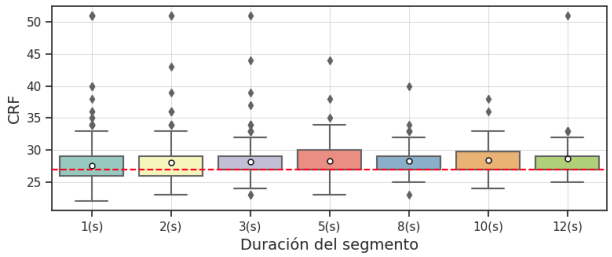
múltiples sistemas operativos e infraestructuras. Se utilizan dos contenedores, el primero implementa el sistema de codificación con sus diversos componentes para la preparación del contenido en múltiples representaciones. El segundo contenedor se utiliza para el análisis de calidad y ofrece soporte para diferentes métricas. El contenedor puede tomar como entrada el vídeo distorsionado y el vídeo de referencia para el análisis de las métricas con referencia completa y referencia reducida; o sólo el vídeo codificado para las métricas sin referencia.

Tabla II: Características del hardware utilizado en las pruebas

Atributo	Detalles
CPU	Intel(R) Core(TM) i7-4790 @3.60GHz
Cache size	8192 KB
Placa Base	ASUS MB H81M-D PLUS
Memoria RAM	Kingston 99U5403-159.A01LF 2x8 GB DIMM DDR3 1600 MHz
Almacenamiento	WD40NDZW-11A8JS1, 5400 RPM
SO	Linux Mint 17.3 Rosa x64 bits
Docker Versión	19.03.5



(a) Vídeo STL



(b) Vídeo TOS.

Fig. 3: Distribución de los CRFs calculados para diferentes tamaños de segmento.

B. *Experimentos y Resultados*

En esta subsección presentamos los experimentos realizados y los resultados obtenidos utilizando el esquema de codificación propuesto. Como hemos indicado en apartados anteriores, disponemos de 2 vídeos en formato Y4M (STL y TOS) con resolución 4K. Cada vídeo ha sido codificado para siete tamaños de segmento (1, 2, 3, 5, 8, 10 y 12 segundos) y para siete resoluciones (240p, 360p, 480p, 720p, 1080p, 1440p y 2160p). Debido a las restricciones de longitud del documento, nos centraremos en la resolución 1080p, pero hay que señalar que los resultados mostrados con esta resolución son similares para las demás resoluciones evaluadas. Nuestra propuesta se ha comparado con una codificación basada en segmentos utilizando un CRF constante.

1) Selección de valores constantes del CRF: Para obtener el valor constante de CRF con el que comparar nuestra propuesta, hemos codificado los vídeos completos utilizando diferentes valores de CRF como se muestra en la Tabla III. Con un CRF de 27 el VMAF obtenido es mayor o igual a 90, que es el valor de calidad objetivo utilizado como entrada en nuestro sistema.

2) Experimentos con diferentes tamaños de segmentos: La distribución de los CRFs para los diferentes tamaños de segmento de los dos vídeos considerados se muestra en la Fig. 3. Como se puede observar, la media es muy cercana a 27, pero el esquema propuesto permite la adaptación al contenido del vídeo. Cabe destacar que estos valores se han obtenido para la dimensión reducida (240p) y se han aplicado al resto de las siete resoluciones indicadas al inicio de esta subsección.

Las tablas IV y V presentan un resumen de los datos obtenidos durante la codificación de diferentes tamaños de segmento con los vídeos STL y TOS, respectivamente. Las tablas muestran, para cada duración de segmento, el número de segmentos, la media aritmética de VMAF, la desviación estándar de VMAF, el tiempo de codificación y el tamaño del vídeo codificado. Estos datos se muestran para una codificación de CRF fijo de 27 y para CRF adaptativo (nuestra propuesta) con un VMAF objetivo de 90. Finalmente, la última columna presenta una comparación entre el tamaño de la codificación CRF fija y adaptable. Un valor negativo significa que la codificación fija mejora nuestra propuesta mientras que un valor positivo significa que nuestra propuesta aporta beneficios.

En la película Sintel (ver Tabla IV), podemos ver que nuestra propuesta tiene un VMAF medio de 92.17, mientras que el VMAF medio de la codificación CRF fija es de 92.79. Esta diferencia de 0.62 en la métrica VMAF no es perceptible para los humanos y menos aún cuando se trata de valores VMAF superiores a 88 (considerados de muy alta calidad). Nuestra propuesta tiene una desviación estándar menor, lo que implica un mejor ajuste a la calidad objetivo.

Si nos fijamos en el tamaño del vídeo codificado, nuestra propuesta mejora en todos los casos. Tenemos un porcentaje de mejora superior al 8.4% cuando los segmentos tienen un tamaño igual o superior a 3 segundos. En términos medios, nuestra propuesta mejora el tamaño del vídeo en un 8.8%.

En el vídeo Tears of Steel (véase la tabla V), podemos ver que nuestra propuesta tiene un VMAF medio de 91.31, mientras que el VMAF medio de la codificación CRF fija es de 92.07. Esta diferencia de 0.76, inferior a 1 y similar a la obtenida con el vídeo STL. Nuestra propuesta tiene una desviación estándar menor, lo que implica un mejor ajuste a la calidad objetivo, en este caso en torno a 90.

Si nos fijamos en el tamaño del vídeo codificado, nuestra propuesta mejora en todos los casos, excepto con segmentos de 1 segundo. Tenemos un porcentaje de mejora

Tabla III: Vídeos codificados con diferentes CRF (sin segmentar)

CRF	STL			TOS		
	VMAF	*Tiempo (min)*	*Tamaño (MB)*	*VMAF*	*Tiempo (min)*	*Tamaño (MB)*
15	98.504	21.035	1554.066	98.810	20.230	2676.059
21	97.057	20.724	642.715	97.005	17.669	724.297
23	96.137	19.145	484.047	95.912	15.951	506.457
25	94.851	18.867	366.625	94.415	15.816	370.617
27	92.151	18.743	281.148	92.376	15.600	280.242
29	90.687	18.685	218.363	89.694	15.534	216.645
31	87.552	18.591	171.367	86.225	15.366	170.145
35	78.657	18.371	109.234	76.791	15.332	108.871

Tabla IV: Comparación de los datos del proceso de codificación para Sintel para la resolución 1080p.

STL		CRF Fijo				CRF Adaptativo				Fijo vs Adaptativo
Seg. Duración	*Seg. Número*	*VMAF*	*Des Est*	*Tiempo (min)*	*Tamaño (MB)*	*VMAF*	*Des Est*	*Tiempo (min)*	*Tamaño (MB)*	*Tamaño (%)*
1(s)	888	92.269	5.085	38.481	321.062	92.326	4.226	34.839	315.59	1.704
2(s)	444	92.657	4.958	36.863	297.793	92.24	4.355	33.238	278.422	6.505
3(s)	296	92.843	4.952	33.163	291.078	92.176	4.479	27.453	266.523	8.436
5(s)	178	92.896	4.857	24.457	284.039	92.158	4.486	25.206	255.891	9.910
8(s)	111	92.896	4.857	23.304	280.727	92.139	4.531	21.76	250.27	10.849
10(s)	89	92.977	4.825	22.532	279.672	92.103	4.629	21.961	247.285	11.580
12(s)	74	92.988	4.818	21.049	282.645	92.108	4.662	23.141	250.184	11.485

Tabla V: Comparación de datos del proceso de codificación para Tears of Steel para la resolución 1080p.

TOS		CRF Fijo				CRF Adaptativo				Fijo vs Adaptativo
Seg. Duración	*Seg. Número*	*VMAF*	*Des Est*	*Tiempo (min)*	*Tamaño (MB)*	*VMAF*	*Des Est*	*Tiempo (min)*	*Tamaño (MB)*	*Tamaño (%)*
1(s)	735	91.457	4.936	31.33	308.406	91.513	4.193	28.772	315.964	-2.451
2(s)	368	91.905	4.798	26.026	292.754	91.363	4.36	26.159	279.816	4.419
3(s)	245	92.052	4.744	25.964	287.355	91.28	4.347	25.493	267.864	6.783
5(s)	147	92.212	4.703	21.815	283.43	91.26	4.507	23.576	256.788	9.400
8(s)	92	92.294	4.666	18.483	281.633	91.313	4.445	18.628	252.936	10.190
10(s)	74	92.304	4.652	19.345	280.676	91.189	4.6	18.961	248.24	11.556
12(s)	62	92.298	4.65	17.983	282.074	91.263	4.587	18.606	248.996	11.727

superior al 6.78% cuando los segmentos tienen un tamaño igual o superior a 3 segundos. En términos medios, nuestra propuesta mejora el tamaño del vídeo en un 7.5%.

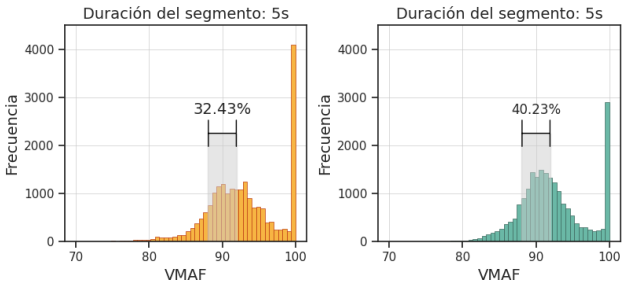
Los tiempos de codificación con CRF fijo y CRF adaptativo son similares. En el caso adaptativo el tiempo mostrado corresponde al tiempo empleado para codificar el video a la resolución de 1080p. El tiempo medio empleado por el Core Loop (que calcula los valores de CRF para cada segmento a la resolución de 240p) es 12,28 min para STL y 16,36 min para TOS. No se suma este tiempo a las tabla IV y V puesto que esta etapa se realiza una sola vez para todo el rango de resoluciones seleccionadas; en nuestro caso las siete resoluciones mencionadas al inicio de está sección.

3) Análisis de calidad para segmentos de 5s: Varias empresas dedicadas a la transmisiones HAS indican que el tamaño de los segmentos puede variar mucho, ya que la calidad percibida por el usuario depende de éste y otros muchos factores. En [11], los autores recomiendan un tamaño de segmento igual o inferior a 6 segundos, por lo que centraremos nuestro estudio en la codificación con segmentos de 5 segundos. Para comprobar si nuestro sistema se ajusta a la calidad objetivo (en este caso VMAF=90) cuando los dos vídeos están codificados a una resolución de 1080p, hemos obtenido la calidad VMAF de cada fotograma. La Fig. 4 muestra la distribución de VMAF para los dos vídeos STL (arriba) y TOS (abajo) y los dos esquemas de codificación: CRF constante (izquierda) y CRF adaptativo (derecha) para un tamaño de segmento de 5s.

La Fig. 4 muestra un porcentaje en cada figura. Este porcentaje corresponde al número de fotogramas que tienen un VMAF entre 88 y 92. En STL vemos que nuestra propuesta tiene un 40.23% de los fotogramas en torno a la calidad objetivo (90), aproximadamente un 7% más que la solución con CRF fijo. En TOS, nuestra propuesta tiene un 42.6% de los fotogramas alrededor de la calidad objetivo, ligeramente superior al 38.78% obtenido con CRF fijo. Sin embargo, vemos que hay más valores de VMAF muy cercanos a 100 en el caso de CRF constante. La Fig. 5 muestra la distribución de VMAF por fotograma para los dos vídeos y esquemas de codificación. Podemos ver que el caso de CRF constante da valores muy cercanos a 100 en los fotogramas finales (créditos finales) en ambos vídeos mientras que nuestra propuesta da valores de calidad ligeramente inferiores.

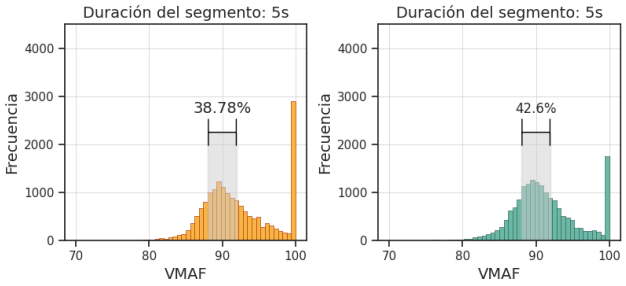
V. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se ha presentado un nuevo esquema de codificación para DASH que es capaz de adaptar el CRF por segmento para lograr una calidad preestablecida. El usuario proporciona el vídeo de entrada en alta res-



(a) CRF 27 - STL.

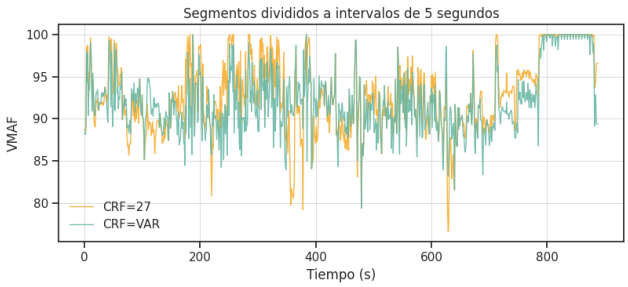
(b) CRF Adaptativo - STL.



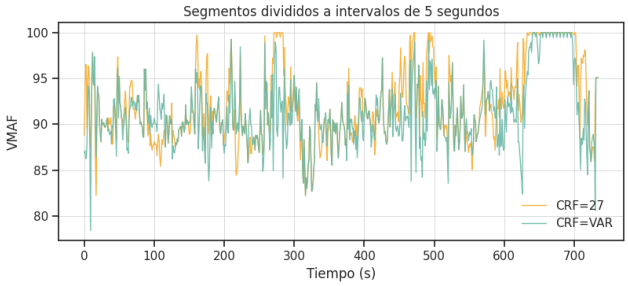
(c) CRF 27 - TOS.

(d) CRF Adaptativo - TOS.

Fig. 4: Distribución VMAF para dos vídeos, STL superior y TOS inferior, con un segmento de 5s.



(a) VMAF por fotograma para STL.



(b) VMAF por fotograma para TOS.

Fig. 5: VMAF por fotograma para los dos vídeos.

olución, la duración del segmento, la calidad objetivo y las resoluciones. El vídeo se reduce a una resolución de 240p, se divide en segmentos y se codifica para diferentes valores de CRF. Se obtiene la calidad de cada segmento codificado y se encuentra el CRF que da la calidad deseada. Estos valores se utilizan para codificar el vídeo original a diferentes resoluciones. Los experimentos con dos vídeos y diferentes tamaños de segmento muestran que este esquema puede reducir el tamaño del vídeo codificado en torno al 10% en comparación con un CRF fijo por segmento.

Como trabajo futuro, compararemos la división uniforme con una división basada en escenas y codificaremos cada segmento en un entorno de nube para aumentar el rendimiento.

Agradecimiento

Este trabajo ha sido apoyado por el programa español MCIU bajo la subvención RTI2018-098156-B-C55 y la red de excelencia RED2018-102383-T. También dicho trabajo ha sido apoyado por la Universitat de València bajo la ayuda UV-INV-AE-1564749.

Referencias

[1] C. Systems, "Cisco annual internet report (2018–2023)," Cisco Systems, White Paper, March 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html

[2] Sandvine, "2020 covid internet phenomena spotlight report," Sandvine Incorporated, Report, May 2020. [Online]. Available: https://www.sandvine.com/covid-internet-spotlight-report

[3] A. C. Begen and Y. Syed, "Are the streamingformat wars over?" in *2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2018, pp. 1–4.

[4] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *The Netflix Tech Blog*, vol. 6, no. 2, 2016.

[5] A. Aaron, Z. Li, M. Manohara, J. De Cock, and D. Ronca, "Per-title encode optimization," *The Netflix Techblog*, 2015.

[6] J. De Cock, Z. Li, M. Manohara, and A. Aaron, "Complexity-based consistent-quality encoding in the cloud," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1484–1488.

[7] Y. Lin, H. Denman, and A. Kokaram, "Multipass encoding for reducing pulsing artifacts in cloud based video transcoding," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 907–911.

[8] H. Xing, Z. Zhou, J. Wang, H. Shen, D. He, and F. Li, "Predicting rate control target through a learning based content adaptive model," in *2019 Picture Coding Symposium (PCS)*, 2019, pp. 1–5.

[9] M. Covell, M. Arjovsky, Y. Lin, and A. Kokaram, "Optimizing transcoder quality targets using a neural network with an embedded bitrate model," in *Visual Information Processing and Communication*, 2016.

[10] H. Wei, X. Zhou, W. Zhou, C. Yan, Z. Duan, and N. Shan, "Visual saliency based perceptual video coding in hevc," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2016, pp. 2547–2550.

[11] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of http adaptive streaming," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Cloud QoX: arquitectura del sistema de recogida de información. Aproximación en Educación.

Rosa Mora, Julián Fernández-Navajas, José Ruiz-Mas, Ana Cebollero, Patricia Chueca, Marta Lampaya.
Departamento de Ingeniería Electrónica y Comunicaciones, Departamento de Ciencias de la Educación

Universidad de Zaragoza - Calle María de Luna, 3, 50018 Zaragoza (Spain).

rosa.mora@zafiro.tv, navajas@unizar.es, jruiz@unizar.es, anacebollero@unizar.es, 738881@unizar.es, 736487@unizar.es

Hace décadas que nuestra Sociedad comenzó a resolver ciertos problemas con herramientas digitales, gracias a las ventajas que aportan productos y servicios digitales a la hora de satisfacer ciertas necesidades. Por esto, la digitalización se ha desplegado de forma progresiva en todos los sectores y actualmente se está acuñando un nuevo paradigma: el bienestar digital, que podemos medir mediante la QoX (Calidad de X), siendo la X cualquier aspecto que se trate sobre el bienestar digital.

En nuestra investigación pretendemos generar una red digital de conocimiento, servicios y productos. Nuestro primer objetivo es proponer una arquitectura funcional y desarrollar herramientas básicas que den soporte a los grupos expertos relacionados con el bienestar digital. Proponemos conjuntamente un modelo que permita dimensionar los problemas, plantear soluciones y gestionar recursos orquestando conocimiento y tecnología. Las sucesivas herramientas apoyarán a las comunidades de expertos a prevenir, anticipar y enfrentar situaciones de riesgo. Explicaremos nuestra propuesta de arquitectura funcional sobre el caso de uso de la situación a que se enfrenta el Sector de la Educación ante ciertos problemas detectados tras el arrollador éxito de los modelos de negocio en servicios multimedia, que son: el elevado contenido emocional en las redes sociales, el multitasking [1], el FoMo [2] [3] (Fear Of Missing Out, o ansiedad tecnológica, miedo a perderse algo), el phubbing [4] [5] (phone-snubbing, o mirar al móvil mientras hay una comunicación interpersonal).

***Palabras Clave*- QoX, bienestar digital, servicios digitales.**

I. INTRODUCCIÓN

La Sociedad de la Información evoluciona durante las últimas décadas de forma espectacular, gracias a definir retos digitales, tanto desde el punto de vista del tecnólogo como del usuario final. Esto ha provocado que temamos que pequeños detalles puedan hacer caer todo el sistema, como sucedió cuando pasamos del año 1999 al 2000 y se temió por la caída generalizada de ordenadores y redes de comunicaciones. Además, el desarrollo de las tecnologías no debe estar reñido con las ideas de aldea global o el compromiso con la sostenibilidad. Durante la última década, hemos trabajado los objetivos 2020 y la visión mundial del 2050. Se dibujan escenarios internacionales y definen nuevos retos digitales, con hitos que lideran las Naciones Unidas, donde caben todos los orígenes, edades y géneros, y se consideran los deseos y temores de forma única [6]: ***"Across the world, respondents of all origins, genders and age groups, are remarkably unified in their fears and hopes for the future"***. En un escenario como éste, la pandemia COVID-19 nos ha cuestionado a todos y refuerza el compromiso de cooperación digital. El informe UN 75 Anniversary nos pide ***"(…) Hard work towards Universal Access to Digital Tec, equitable shift to digital and online education"***. La digitalización de los procesos nos lleva a universalizar servicios que antes únicamente estaban disponibles en ciertas localizaciones y para ciertas franjas horarias. Debemos definir y desarrollar arquitecturas funcionales que den una respuesta global y estén disponibles 24 horas, 7 días a la semana. Debemos ofrecer monitorización, análisis de resultados on-line y facilitar la actuación en caso de necesidad.

Fig. 1. DQ Institute Global Standards (dqinstitute.org)

En la Figura 1 mostramos cómo el DQ Institute clasifica el nuevo paradigma digital y define estándares globales, como Seguridad Digital, Derechos Digitales, Inteligencia Emocional Digital, etc. Por tanto, definimos nuestra propuesta de arquitectura funcional hacia un modelo de desarrollo sostenible. La arquitectura debe soportar diferentes sectores y mercados, debe contribuir al desarrollo de una base tecnológica de aplicaciones y estándares globales orientados al crecimiento sostenible.

Además, cuando la tecnología de comunicaciones evolucionó desde broadcast hacia multicast/unicast, se regularon aspectos éticos en las normativas nacionales. En España encontramos referencias claras de la Comisión del Mercado de Telecomunicaciones, desde abril de 2010 con el BOE-A-2010-5292 de la Ley Audiovisual, hasta mayo de 2014 con el BOE-A-2014-4950 de la Ley General de Telecomunicaciones, o el Plan de Actuación CNMC de febrero de 2019. Es decir, un nuevo planteamiento técnico, una nueva arquitectura funcional conlleva necesariamente responder a los retos humanísticos de siempre. Estamos ante una evolución natural donde la comunicación masiva podría ser multicast y principalmente sobre dispositivos multimedia. Hay que encontrar una solución de compromiso entre privacidad y perfilado de usuario, modelos de negocio y calidad de experiencia [7] que garanticen la adecuación del servicio. Hay que encontrar el punto adecuado para monitorizar la calidad del servicio, la seguridad y satisfacción del usuario. Constatamos cómo, en los últimos años, miles de millones de usuarios acceden a redes sociales, tal como ilustra la Figura 2; algo que no pasaba a finales de los años 90. Los modelos de negocio asociados a estas redes tienen un impacto personal, que debe tenerse en cuenta, y no sólo quedarnos con el beneficio empresarial o de conectividad. Por ello, nos planteamos qué medios técnicos deberíamos desarrollar para analizar, convenientemente, esta nueva situación.

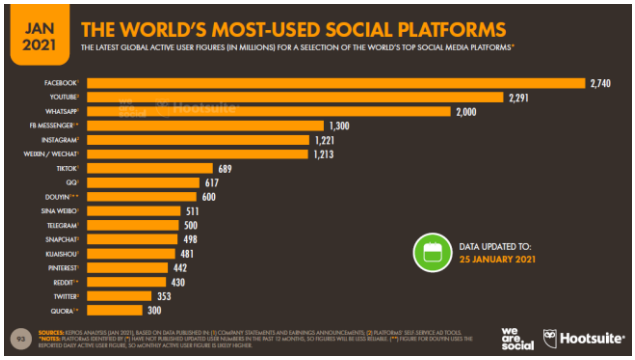


Fig. 2. Redes Sociales y objetivos de penetración de banda ancha: de cero a miles de millones de usuarios en sólo dos décadas [8].

Muchas comunidades de expertos ya han encontrado cómo identificar riesgos, y transformar conductas en competencias. Nos disponemos a analizar el caso concreto de la comunidad de educadores, y cómo la red DQ Institute define una matriz que les permite comprender la transformación digital que experimenta el individuo y el grupo donde se mueve [9]. Esta organización aúna esfuerzos públicos y privados y estudia aspectos tales como identidad, uso, derechos, comunicación, seguridad, inteligencia emocional, etc. También define todo un ecosistema de funcionalidades y servicios que proporcionan a su vez las empresas y organizaciones que la constituyen, como líderes en su sector. En la Figura 3 mostramos el modelo que han desarrollado, de manera muy completa y minuciosa para comprender el paso de conductas a competencias, capaz de relacionar habilidades (como la empatía o la identidad digital del ciudadano), con fortalezas psicológicas (como el pensamiento crítico y el autocontrol), con el comportamiento para prevenir el riesgo cibernético (como la prevención de la adicción tecnológica, la intrusión, o el ciberacoso), y con el desarrollo cognoscitivo y social (rendimiento académico, ciudadanía global).

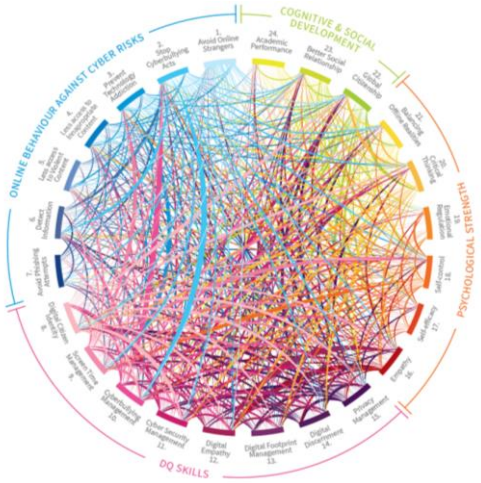


Fig. 3. DQ Institute: de conductas a competencias (dqinstitute.org)

Por otra parte, diversas investigaciones en psiquiatría y psicología advierten que el abuso de los servicios digitales podría desembocar en problemas personales, tales como la adicción a Internet o el stress digital [10][11]. Si bien, ya estaban identificados como problemas en el entorno laboral, en la actualidad afecta también al entorno personal. Vemos cómo ciertos gobiernos locales proponen a sus ciudadanos programas de desintoxicación digital y las empresas, cursos de gestión eficaz del tiempo como medida de choque frente a estos problemas. Sin embargo, cada vez se tiene más consciencia sobre la necesidad de programas de formación personal para abordar los problemas desde edades tempranas.

En resumen, no sólo debemos desarrollar tecnología pensando en la Calidad de Servicio (QoS Quality of Service), o la Calidad de Experiencia (QoE, Quality of Experience) sino en la Calidad de Cualquier Incógnita que surja (QoX), muy especialmente las relacionadas con las emociones y el estrés. En este artículo se pretende definir una arquitectura funcional QoX que responda al reto del estrés digital. En primer lugar, presentaremos los antecedentes del problema, la propuesta de diseño, implementación y conclusiones preliminares de dicha arquitectura funcional QoX.

II. ANTECEDENTES: GRUPOS DE TRABAJO Y OBJETIVOS

Tal y como planteamos en la Introducción, estamos centrando nuestro diseño de arquitectura QoX en los requerimientos que actualmente tienen los programas de formación y educación. Dichos programas, para ser realmente efectivos, deben basarse en un análisis correcto de las diferentes situaciones de riesgo. Para eso se hace necesario el desarrollo de una herramienta que recoja datos sobre la utilización de las nuevas tecnologías. La recolección de los datos deberá hacerse de la forma más objetiva posible. Todo ello se fundamentará en la estrecha colaboración entre un grupo de expertos en educación y otro en tecnologías digitales. La herramienta debe ayudar a captar los problemas y necesidades de los usuarios de los

servicios digitales e incluso, en un futuro, utilizarse como soporte para estimular el uso responsable, gracias a una ingeniería sostenible de comunicaciones.

Hemos colaborado dos grupos de trabajo. El grupo EDUCAVIVA (expertos en educación y psicología) y el grupo CeNIT (expertos en TICs). Las aportaciones de estos dos grupos las resumimos a continuación.

El grupo EDUCAVIVA dispone de un valioso conocimiento previo, referencias, una red de profesionales en marcha y grupos de personas donde ya se han experimentado y medido los efectos de la digitalización. Consideran una serie de criterios de desarrollo, para enfrentarse a los potenciales problemas derivados / detectados durante el proceso de digitalización y penetración masiva de las redes sociales. A su vez, estos problemas son inherentes al crecimiento exponencial en el tráfico, número de usuarios y penetración de los servicios en banda ancha y movilidad. Crecimiento que se ha visto potenciado por los propios agentes (proveedores de servicios y contenidos), que maximizan su beneficio al fomentar el consumo.

Veamos a continuación unos análisis iniciales de datos que manejan junto con la bibliografía científica [12]-[31], ya que parece que el uso problemático de internet (o abuso) tiene una alta correlación con una serie de factores:

1. Alto contenido emocional en línea: que incluye la expresión emocional que hacemos en las redes sociales (e.j. expresar si se está triste o alegre) y la facilitación a los demás de las propias emociones (expresarlas para mejorar las relaciones con los contactos, para que se sientan comprendidos, para superar dificultades, etc.). Aunque se tenga una competencia socioemocional desarrollada en general, el contenido emocional en línea disminuye en mucho su factor de protección ante la adicción a las redes sociales [12]. Por tanto, sería necesario cuantificar ese contenido emocional en línea (e.j. conteo de emoticonos que emite y recibe una persona).

2. FOMO (ansiedad de perderse algo): que se define como la percepción generalizada de que otros puedan estar viviendo experiencias gratificantes de las cuales uno está ausente [13]. FOMO es un fuerte predictor del abuso de internet y de las redes sociales [14]. Este indicador está relacionado con otros como: la impulsividad, la necesidad de reconocimiento personal por parte de otros, tiempo dedicado a la imagen personal y a las relaciones sociales, todos ellos factores relacionados con el abuso a Internet. Se podría cuantificar el FOMO a través del número de desbloqueos del móvil, el número de accesos a las redes sociales habituales y tiempo que se está en ellas, tiempo que se tarda en contestar a mensajes recibidos en las redes sociales, tiempo de imagen personal (e.j. cambio de perfil y estado), número de publicaciones de fotos o videos.

3. Multitarea: se consideran dos tipos, el uso de varias aplicaciones y dispositivos tecnológicos al mismo tiempo y el uso de la tecnología mientras se realiza una actividad no multimedia [15]. En adolescentes, la multitarea es especialmente intensa [16]. Las investigaciones indican que un mayor uso de la tecnología provoca un aumento de la multitarea y ésta parece ser una razón por la que se hace phubbing y aumenta el abuso de la tecnología [17]. Tiene numerosos efectos cognitivos y afectivos: un peor funcionamiento cognitivo en tareas que implicaban memoria de trabajo y velocidad de procesamiento e informaron de menor rendimiento. También indican una peor atención y regulación emocional [18] [19] [20] [21]. Los grupos con elevados niveles de multitarea de medios usan emoticonos con más frecuencia, usan emoticonos en múltiples dispositivos y actualizan los emoticones con más frecuencia [22]. Para cuantificar la multitarea podría monitorizarse los minutos de uso seguido en una APP, y el número de APPs utilizadas al mismo tiempo.

4. Phubbing: es el acto de mirar al móvil mientras se conversa cara a cara con otras personas. Se ha convertido en una rutina comunicativa con un gran impacto en las personas: disminuye la calidad de la comunicación interpersonal [23], la satisfacción de las relaciones [24], aumenta la sensación de ser devaluado para quien los sufre [25] aumento de sentimiento de celos [26], de falta de intimidad con los socios [27] e incluso niveles más altos de depresión [28], ansiedad social y emocional [29]. El phubbing es una causa y efecto de la adicción al móvil, a internet y a las redes sociales. Además de promover la reducción de la adicción, puede gestionarse con la consciencia de la misma y de otras acciones en internet, ya que provoca autocontrol en el comportamiento y decrecen otras adicciones [30], como la adicción a los teléfonos inteligentes, los SMS y los medios de comunicación [31]. Podríamos cuantificar este factor si el dispositivo móvil tuviera los sensores adecuados para detectar si se usa mientras se mantiene una conversación.

EDUCAVIVA ha analizado en detalle el caso de las aplicaciones que usan los menores, como por ejemplo WhatsApp, Instagram o Facebook y se ha planteado la posibilidad de desarrollar una solución técnica de compromiso que tendría que considerar en su etapa de especificación del desarrollo no sólo los criterios del proveedor de servicios, sino también una serie de criterios educativos, para minimizar los problemas de adicción a internet ya desde la etapa de diseño de una herramienta con conexión a internet.

Por otro lado, el grupo CeNIT (Communication Networks and Information Technologies group), ha analizado una serie de trabajos sobre los nuevos modelos de servicio del despliegue de la tecnología IPTV, y sus medidas de calidad asociadas [32][33]. Esto supone un paso previo que puede ser adaptado al resto de tecnologías de comunicación digital. Ha planteado en QQCM el desarrollo de un modelo de calidad [34], considerando la evolución QoS (Quality of Service) y QoE (Quality of Experience), hasta QoX (cualquier nuevo concepto que surja al profundizar en la percepción del usuario final). En definitiva, trabaja en el desarrollo de arquitecturas funcionales que deben abarcar redes y nuevas tecnologías, compartir conocimiento, generar experiencias y sobre todo, posibilitar la planificación de actividades a diferentes niveles funcionales de forma ordenada.

Además, CENIT, ha estudiado también distintas aplicaciones desarrolladas con la finalidad de planificar el tiempo de bloqueo de aplicaciones y llamadas como (OFFTIME), Quality Time, Screen Time, Rescue Time,

Flipd, Forest, Space, Qustodio,D Drive Mode, Waze, etc. De alguna forma, permiten personalizar el tiempo, apoyando al usuario en su proceso personal de desintoxicación digital. Claro, sólo para aquellos usuarios que ya son conscientes de que precisan desconectarse y disfrutar de un tiempo de calidad.

Se plantea, por tanto, el reto tecnológico de forma similar a otros despliegues de tecnologías, basándonos en teorías matemáticas como la de vectores y valores propios para identificar las componentes más relevantes. Y en la teoría de antenas y su diagrama de radiación como forma gráfica de expresión, para poner de relieve esos conflictos de intereses y sus soluciones de compromiso. Ambos modelos teóricos posibilitan una nueva generación TIC, y vienen avaladas por el éxito en despliegues de antenas en topologías complejas, problemas en los modelos de negocio, restricciones en el acceso a la energía, pérdidas y daños por disipación, etc.

Para dar una solución tecnológica a este reto, partimos hace ya varias décadas de los problemas propios de la pérdida de paquetes en comunicaciones, la escasez de recursos a la hora de transmitir información [35][36] y cómo desplegar sistemas que tuvieran en cuenta la percepción del usuario: lo que era realmente capaz de apreciar y lo que podría considerar no relevante, o reconstruirse a posteriori [37]. Como resumen, la Figura 4 muestra la propuesta de EHU (Euskal Herriko Unibertsitatea) para cuantificar la QoE (Calidad de Experiencia) hacia la QoX [38].

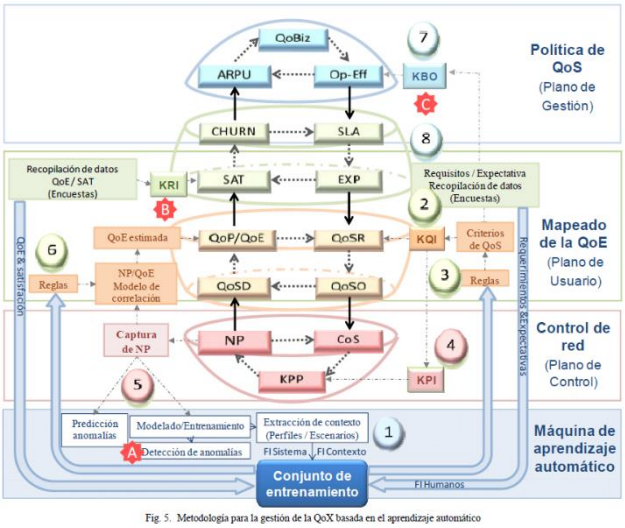


Fig. 5. Metodología para la gestión de la QoX basada en el aprendizaje automático

Fig. 4. Propuesta QoE a QoX del grupo de trabajo EHU.

Para establecer de forma correcta la cooperación entre sendos grupos, nos planteamos hacer frente a este nuevo problema de salud digital en la realidad de su día a día, analizando la percepción humana y el comportamiento en ciertos grupos de personas. Para ello, el trabajo conjunto debe enfocarse en la monitorización y actuación sobre el elevado contenido emocional en las redes, o diferentes situaciones de riesgo como son la multitarea, el FoMo (fear of missing out) y el Phubbing (phone-snubbing). Con los resultados obtenidos y posteriormente analizados, estaríamos en disposición de proporcionar una primera aproximación a los requerimientos que precisan los tecnólogos para generar y definir un modelo digital de referencia.

Esta cooperación debe permitir desarrollar un modelo de QoX para la percepción humana, como una arquitectura funcional cercana a los modelos de calidad propuestos para las redes de comunicaciones, las Smart Grids y sus medidas de calidad y servicio [39] a [46]. Los resultados obtenidos al poner en marcha el concepto con las primeras versiones beta, se analizarán gracias a los modelos teórico-prácticos y estadísticas asociadas. Analizaremos las comunicaciones interpersonales y las acciones que se emprenden después. Compararemos con el modelo ideal, basándonos en el comportamiento aceptado. Definiremos un plan de acción y mejoras sucesivas, propondremos indicadores y medidas (KPI - key performance indicators).

El objetivo final del modelo QoX es ayudar a reorganizar las prioridades de usuarios, usuarios-expertos y sectores mediante una orquestación automática de los recursos disponibles. Por tanto, se nos plantea como requerimiento del sistema un modelo QoX con una arquitectura funcional, que sea lo suficientemente genérico como para replicarlo con usuarios de diferentes grupos / sectores, con diferentes criterios y hasta en conflicto de intereses (por ejemplo, intereses educativos frente a intereses de marketing). Como criterio de diseño, el modelo ha de permitir a cada grupo experto definir sus propios objetivos y métricas, ser flexible para reflejar las necesidades de expertos y usuarios, reflejar cómo se asignan los recursos y cómo se gestionan los conflictos.

Desde el punto de vista humano, parece que estamos de nuevo ante el mito de la caverna que propuso Platón hace más de 2.300 años. Estamos intentando comprender lo que realmente hay detrás de nuestra percepción y nuestras necesidades. Si responden a una realidad absoluta e inamovible, o por el contrario, son percepciones personales que debemos racionalizar, priorizar y gestionar para orientarnos hacia modelos de desarrollo sostenible.

Resumiendo, al analizar los cuatro factores definidos por EDUCAVIVA y su posible cuantificación, nos preguntamos si no estamos ante la necesidad de desarrollar una inteligencia artificial para la adecuada evolución del entorno digital saludable per se. Hemos avanzado en comprender estos criterios y trasladarlos al entorno técnico como una serie de requisitos y arquitectura funcional, tanto para el Front End que atiende a los usuarios (móviles, PCs, etc), como para el Back End que soporta accesos y servicios.

III. DISEÑO DE LA ARQUITECTURA CLOUD QoX

En esta sección proponemos una arquitectura funcional Cloud QoX que pone el acento en el usuario final. Nuestro diseño (ver Figura 5) sigue el modelo EHU sobre QoE mencionado anteriormente, replicando dicho modelo para diversos sectores con criterios dispares, en potencial conflicto de intereses. Nuestro diseño desarrolla una metodología de gestión QoX basada en aprendizaje automático, donde el propio diseño de herramientas digitales haga consciente al usuario de la necesidad de desconexión digital y proporcione un índice de su bienestar digital (contrario al abuso o adicción). Así, partiendo de un

cuestionario inicial o percepción personal inicial, avanzamos de forma comparativa según la evolución de cada persona y también con la media de los grupos a los que un usuario pudiera pertenecer (Sectores). Cada Sector puede definir sus prioridades, los criterios con los que se organiza. Al expresar estos criterios en forma de ejes, el modelo Cloud QoX (Fig 5) está en posición de organizar los recursos disponibles y asignarlos de forma dinámica en función de los objetivos a corto, medio y largo plazo. Los datos que fluyen son accesibles a los grupos expertos que dan soporte a los criterios y el perfilado de usuarios.

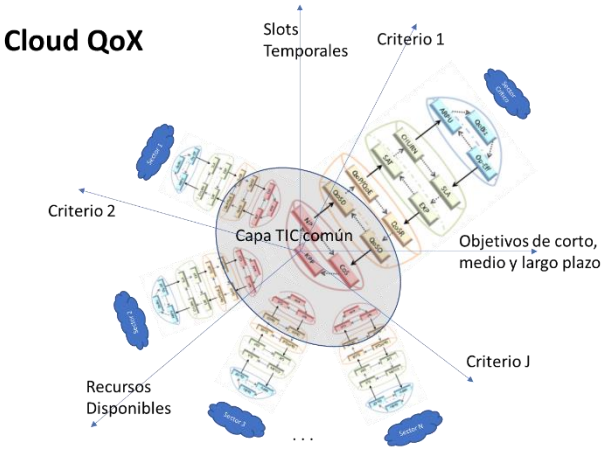


Fig. 5. Modelo Cloud QoX de CeNIT, donde se analizan, priorizan y gestionan objetivos y criterios. Cada Sector o grupo de interés podría tener su propio modelo QoE, recursos y prioridades.

Otra forma de plantearnos el despliegue del modelo QoX, es pensar en los recursos sobre los que se despliega. Planteamos tres grandes grupos (Figura 6). En primer lugar, expertos y usuarios que interactúan gracias a sus emociones y conocimiento. A continuación, dispositivos multimedia, redes de acceso, servidores y servicios en la nube que han de garantizar el adecuado despliegue de una nueva cultura digital. Y por último, una nueva generación de herramientas de análisis que nos ayuden a comprender la realidad y orquestar soluciones para maximizar el bien común. Así, la arquitectura funcional que desarrolla nuestro modelo de QoX debe ser un sistema preparado para ofrecer parámetros monitorizables de diferentes dispositivos multimedia a servicios relacionados con la medida de la percepción.

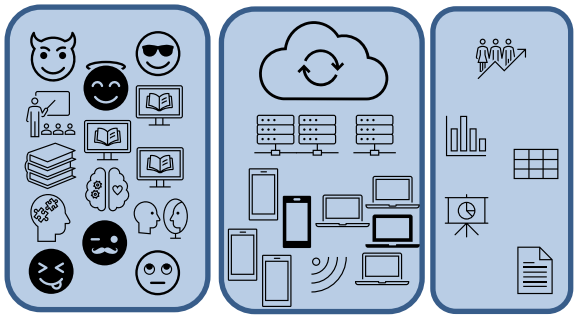


Fig. 6. El modelo de QoX se despliega sobre un grupo de expertos multidisciplinares, con unos entregables definidos. Los dispositivos multimedia serán un factor clave en el análisis de escenarios y planteamiento de soluciones.

Reorganizamos dicha arquitectura funcional para identificar más fácilmente los servicios necesarios para resolver los conflictos de prioridades y criterios de salud digital vs modelo de negocio, tiempo del usuario vs dinero del aplicativo. También debe recomendar a los agentes interesados que mejor puedan dar respuesta a los retos planteados a corto y medio plazo. El modelo permite ir añadiendo servicios y agentes de otros sectores, enriqueciendo la estructura funcional al introducir nuevos agentes, criterios, recursos para cubrir nuevas necesidades.

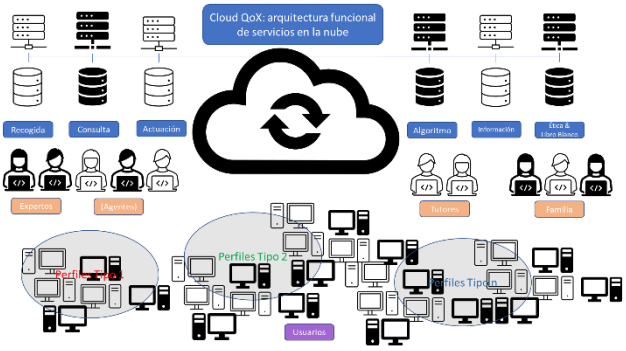


Fig. 7. Arquitectura funcional QoX: servidores y agentes, usuarios y terminales, perfilado de usuario.

Es decir, apostamos por un modelo Cloud QoX que cuenta con:

1.- SERVIDORES Y AGENTES que recopilan la información de los usuarios, permiten al grupo experto la consulta de datos recopilados, disponen la algoritmia para procesado de datos y tratamiento experto con criterios basados en el consenso internacional, gestionan la actuación del experto sobre los terminales de un determinado grupo y permiten la comunicación con el experto. Además, introducimos la figura del servidor donde se recopilan los criterios éticos y se determina el contorno para la comunidad de desarrolladores: se trata de un servidor que contendría los prolegómenos a todo desarrollo futuro, las recomendaciones y guías de buen uso, como evolución del Libro Blanco Digital, orientado a tanto a los desarrolladores de bajo nivel, como UX/UI y los propios gestores de contenido.

2.- TERMINALES Y USUARIOS: los usuarios proporcionan las características que les definen cuando dan de alta sus terminales en el servicio. Estas características permiten al grupo experto la clasificación por grupos. Por tanto, el adecuado perfilado usuario pasa a ser una labor de gran importancia para el experto que monitoriza la calidad del servicio, calidad de respuesta, calidad de experiencia, y define pautas de actuación en caso de necesidad.

3.- REDES DE ACCESO: las distintas tipologías de redes de comunicaciones deben ser analizadas para proponer mejoras en sus despliegues que ayuden a la evolución de las personas, independientemente de la ubicación en la que se encuentren. Por ejemplo, si se detectara un comportamiento diferente en la población rural vs urbana, el equipo experto trataría de compensar las carencias con servicios y funcionalidades adicionales,

hasta que el despliegue físico pudiera habilitar los recursos de red necesarios.

En cuanto a la arquitectura de red, vamos a trabajar sobre el modelo actual, que consideramos válido para soportar la arquitectura funcional QoX, ya que actualmente ya engloba dispositivos multimedia, redes fijas e inalámbricas y servicios en la nube. Además, abordaremos la seguridad como una línea futura, como un factor fundamental de cualquier modelo TIC que se despliegue sobre una población real.

IV. Implementación del sistema para educación

La implementación de nuestro sistema Cloud QoX se ha llevado a cabo como una prueba de concepto para la comunidad educativa. En esta prueba un conjunto de servidores recoge y trata datos obtenidos de un grupo de usuarios de móviles gracias a las herramientas de monitorización de actividad instaladas en los mismos. Tras ello, se analizan los indicadores de ciertos comportamientos para alertar al experto y determinar actuaciones concretas para el individuo o el grupo.

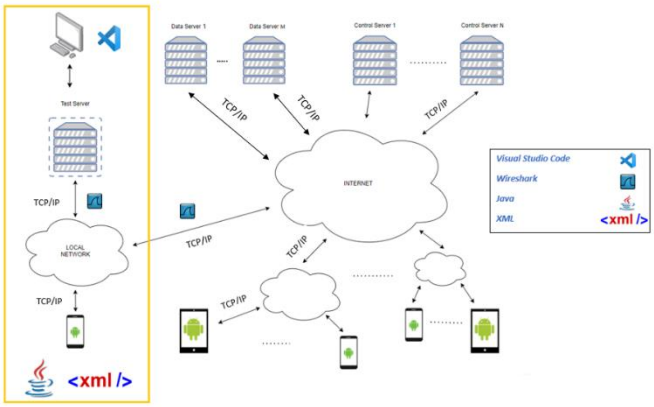


Fig. 8. Implementación de la prueba de concepto Cloud QoX, vista desde el terminal móvil.

La Figura 8 describe el entorno de Cloud QoX visto desde el lado de los terminales móviles. Se ha programado mediante la plataforma Android Studio con soporte para los lenguajes Java y Kotlin y considerando el emulador de Pixel 2 con SO Android 11.0 (API 30), y XML para la UI. Para la comunicación de terminal a servidor se utilizan los protocolos HTTP y COAP. El editor del código fuente usado es MS Visual Studio Code, y se utiliza Wireshark como software para captura y análisis de tráfico de red.

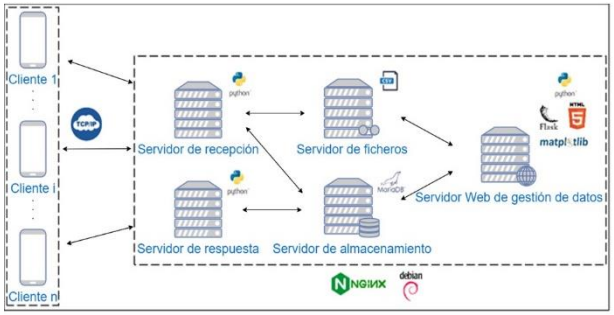


Fig. 9. Implementación de la prueba de concepto Cloud QoX vista desde la infraestructura de servidores.

La Figura 9 muestra la arquitectura funcional de la infraestructura de servidores desarrollada. Cuenta con un sistema operativo Debian; base de datos MariaDB e interfaz gráfica DBeaver; servidor web como proxy inverso NGINX, uWSGI y Certbot. Se ha utilizado Python como lenguaje programación para los servidores con entorno de desarrollo Pycharm, La biblioteca Flask permite crear aplicaciones web, librería de representación de gráficas Matplotlib.y librería de conexión MariaDB; HTML y CSS como lenguaje en el servidor web; exportación de ficheros en formato CSV para los datos en bruto; protocolo de comunicación por sockets TCP.

La actividad del usuario en su terminal móvil, monitorizada con las herramientas comentadas, se ha enviado como eventos a la infraestructura de servidores desarrollada. Se han definido una serie de mensajes como Bloqueo/desbloqueo, encendido/apagado de pantalla, empleo de redes sociales, etc. Que son parte de esa actividad monitorizada y que es trasladada al punto de entrada de la infraestructura (servidor de recepción) vía un conjunto de mensajes (registro, inicio de sesión, datos, actualización de datos, control). Sobre este esquema, en las líneas futuras podrían definirse otros tipos de actividad del usuario y obtenerlos a partir del estudio de los datos.

El servidor de recepción se comunica con el servidor de ficheros y almacenamiento para registrar los mensajes recibidos. Esto posibilitara la generación de ficheros raw con los que los expertos podrían llegar a hacer un procesado de los datos independiente.

El servidor de respuesta proporciona realimentación personalizada a los clientes con los datos existentes en el servidor de almacenamiento, permitiendo al usuario conocer su conducta respecto a la utilización de su dispositivo móvil. Nuestra idea es detectar, informar, alertar y actuar de forma controlada ante potenciales situaciones de riesgo, o prevenirlas antes que se produzcan. La Figura 10 muestra cómo el modelo Cloud QoX hace llegar una alerta a un móvil ante un determinado evento.

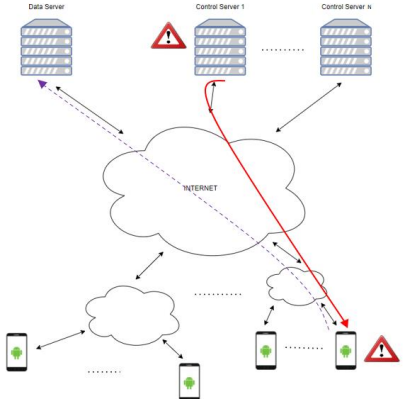


Fig. 10. Retroalimentación personalizada vía servidor de respuesta

Por último, el servidor web de gestión de datos se conecta a los servidores de ficheros y almacenamiento para proporcionar a los expertos una herramienta de acceso a la información recogida y procesada por dicho sistema de servidores. Este tratamiento de la información recogida posibilita precisamente obtener desde aquí la actividad de usuario relacionada con la multitarea o el phubbing.

La Figura 11 muestra un ejemplo de la herramienta gráfica desarrollada para analizar el comportamiento de los distintos grupos. Gracias a la teoría de valores y vectores propios, pensamos que el modelo gráfico sería más útil para el grupo experto que el análisis de datos en bruto. La herramienta les permite comparar el comportamiento del individuo y del grupo frente a sus propios criterios, y ofrecer nuevas recomendaciones para mejorar el modelo teórico. Y para las líneas futuras, hemos preparado ficheros exportables que permitan a los expertos desarrollar sus propios modelos estadísticos sobre las bases de datos en bruto, y anonimadas por seguridad.

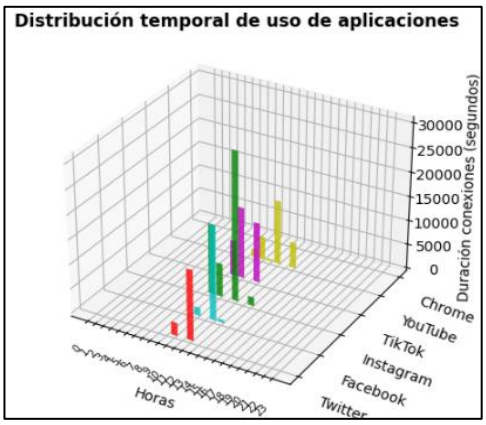


Fig. 11. Una de las gráficas que actualmente muestra el modelo QoX, como presentación de datos de usuario.

V. Conclusiones y lineas futuras

El trabajo aquí presentado define un modelo de arquitectura funcional QoX que responde al reto de estrés digital. Su finalidad es llegar a desarrollar herramientas básicas que den soporte a los grupos expertos relacionados con el bienestar digital para prevenir, anticipar y enfrentar situaciones de riesgo. Estas herramientas ayudarían al experto a identificar y gestionar problemas en sus etapas iniciales, pudiendo prevenir crisis en el individuo o que estas se extiendan al grupo. La propuesta de arquitectura funcional ha sido aplicada al ámbito educativo con la idea de hacer frente a los problemas detectados en el uso / abuso de los nuevos servicios multimedia. Para ello hemos partido de una serie de parámetros técnicos que serían deseables para cuantificar el contenido emocional en las redes sociales (detección de emojis), el multitasking [1] (número de conexiones concurrentes en distintos dispositivos y en el mismo dispositivo, si la tarea desarrollada requiere o no concurrencia), el FoMo [2] [3] (número de desbloqueos del móvil, de entradas en redes sociales y tiempo en ellas, tiempo en acceder y/o borrar notificaciones, número de publicaciones y de qué tipo se trata) y el phubbing [4] [5] (geolocalización, dispositivos cercanos, tipo de conexiones y su frecuencia, interacción con otros usuarios y probabilidad de interacción, intervalos de sueño).

El resultado ha sido la creación de una infraestructura cliente / servidor que permite la recogida y tratamiento de datos obtenidos de un grupo de usuarios de terminales móviles gracias a las herramientas de monitorización de actividad instaladas en los mismos. El posterior análisis de indicadores de ciertos comportamientos permite alertar al experto y determinar actuaciones concretas para el individuo o el grupo. La puesta en marcha de una prueba con usuarios reales es el futuro paso a realizar. Nos permitiría poner a prueba nuestra propuesta de arquitectura funcional y afinar con expertos las posibilidades del trabajo desarrollado. Asimismo, una vez depurado nuestro modelo, se extrapolaría permitiendo modelar la disciplina de trabajo de los usuarios-expertos de los distintos sectores con los que se trabaje, de forma que vayan aportando su percepción y criterios a priori, para evolucionar y mejorar el modelo con sus resultados.

En cuanto a los aspectos de seguridad, proponemos que la securización de comunicaciones y datos extremo a extremo no esté reñida con la seguridad y bienestar digital. Por tanto, debemos continuar trabajando y considerar las lecciones aprendidas en otros casos (como las Smart Grid, etc), y registrar dispositivos para conseguir agregados y estadísticos fiables, así como detección temprana de desviaciones que posibiliten la actuación del experto (Trusted Third Party).

La idea futura es trabajar con nuevos casos reales para analizar cómo la arquitectura funcional QoX en la nube podría dotar de una mejor salud digital a los usuarios. Se precisan nuevos pasos, una discusión multidisciplinar sobre procedimientos y resultados, se trate del ámbito educativo o cualquier otro. Hemos de aprender cómo proteger o crear nuevos entornos digitales, donde no sólo hay que tener en cuenta las tendencias del mercado actual y sus beneficios, sino también el componente ético y de largo plazo. El aprendizaje adquirido se debe recoger en forma de recomendaciones y pautas de buen diseño, un formato de Libro Blanco Digital para también reflejar aquellos desarrollos que no se hayan podido implementar, así como las limitaciones actuales del modelo o de redes, sistemas y dispositivos que lo han impedido. También se reflejarán las ideas más importantes en cuanto a redes, desarrollos, metodología, muestreo, marcos regulatorios, etc. El objetivo es cerrar la actual propuesta con una guía para nuevas versiones de herramientas, procesos y líneas de investigación.

AGRADECIMIENTOS

A los grupos de investigación CeNIT y EDUCAVIVA. Este trabajo ha sido parcialmente financiado por los proyectos T31_20R y S57_20R del Gobierno de Aragón y RED2018- 102383-T del Ministerio de Ciencia, Innovación y Universidades – Agencia Estatal de Investigación.

REFERENCIAS

[1] Martín, Viñas, Malo (2019). Media multitasking impact in homework, executive function and academic performance in Spanish adolescents. Psicothema 2019, Vol. 31, No. 1, 81-87.

[2] Błachnio, Przepiórka (2018). Facebook intrusion, fear of missing out, narcissism, and life satisfaction: A cross-sectional study. Psychiatry Research – Vol 259, Jan 2018, Pages 514–519.

[3] Rosen, Carrier et al (2017). The Role of FOMO in College Course Performance as Mediated by Techn. Usage and Multitasking Habits. Psicología Educativa 2018 V24 N1, P14-25.

[4] Schuur, Baumgartner et al (2017) Media multitasking and sleep problems: A longitudinal study among adolescents. Computers in Human Behavior - Vol 81, Year 2018, P 316 – 324.

[5] Ho Moon, E. Lee, J. Lee, Choi, Sung (2016). The role of narcissism in self promotion on Instagram. Personality and Individual Differences - Volume 101, Oct 2016, Pag 22-25.

[6] United Nations UN 75 Anniversary Report, available at https://www.un.org/en/un75/presskit, last access February 2021.

[7] Martinez; Nesse et al (2015) QoE-based service differentiation: Business models analysis for the mobile market. 26th European Regional Conference of ITS, 24-27 June 2015.

[8] Juan Mejía,Transformación Digital, available at https://juancmejia.com, last access MAY 2021.

[9] DQ Institute Glonal Standard for Digital Literacy and Skills, by the Coalition for Digital Intelligence at https://www.dqinstitute.org/, last Access FEB 2021.

[10] Jerald J. Block M.D. (2008). Issues for DSM-V: Internet Addiction. American Journal of Psychiatry 165:3, March 2008.

[11] Salanova, Llorens, Cifre (2012). The dark side of techn.: technostress among ICT users. International Journal of Psychology. 2013 DOI 10.1080/00207594.2012.680460.

[12] Nasaescu, Marín, Llorent, Ortega, Zych (2018). Abuse of technology in adolescence and its relation to social and emotional competencies, emotions in online comm. and bullying. Computers in Human Behavior - Vol 88, Nov 2018, P114-120.

[13] Przybylskia et al (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behaviour - Vol 29, Iss 4, July 2013, P 1841-1848.

[14] H. Fuster, A. Chamarro, U. Oberst (2018). Fear of Missing Out, online social networking and mobile phone addiction (…). Aloma Vol. 35 Núm. 1 (2017).

[15] Schuur, Baumgartner, Sumter, Valkenburg (2015). The consequences of media multitasking for youth: A review. Computers in Human Behavior, Vol 53, 204–215.

[16] Rideout, Victoria J.; Foehr, Ulla G.; Roberts, Donald F. (2010). Generation M2: Media in the Lives of 8- to 18-Year-Olds. Henry J. Kaiser Family Found., ED527859 - JAN 2010.

[17] Voorveld & van der Goot (2013). Media multitasking across age groups: A diary study. Journal of Broadcasting & Electronic Media July 2013 Vol 57(3) Pages 392-408.

[18] Cain., Leonard, et al. (2016). Media multitasking in adolescence. Psychon Bull Rev 23, 1932–1941 (2016).

[19] Courage, Bakhtiar, Fitzpatrick, et al (2015). Growing up multitasking: The costs and benefits for cognitive development. Developmental Review Volume 35, March 2015, Pages 5-41.

[20] Murphy, McLauchlan, Lee (2017). Is there a link between media-multitasking and the executive functions of filtering and response inhibition? Computers in Human Behavior. Vol 75, Oct 2017, Pags 667-677.

[21] Schuur, Baumgartner et al (2015). The consequences of media multitasking for youth: A review. Computers in Human Behavior. Volume 53, December 2015, Pages 204-215.

[22] L.Liu, Cheng, X.Liu (2018). Research on the Impact of Media Multitasking on Emoticons Usage. ICIME 2018. IEEE Catalog Num: CFP1841G-POD.

[23] McDaniel, Coyne. (2014). "Technoference": The interference of technology in couple relationships and implications for women's personal and relational well-being. Psychology of Popular Media Culture. 5(1), 85–98.

[24] V. Chotpitayasunondh & K. Douglas (2018). The effects of "phubbing" on social interaction. Journal of Applied Social Psychology January 2018.

[25] Abeele, Postma-Nilsenova (2018). More Than Just Gaze: An Experimental Vignette Study Examining How Phone-Gazing and Newspaper-Gazing and Phubbing-While-Speaking and Phubbing-While-Listening Compare in Their Effect on Affiliation. Comm Research Reports. Volume 35, 2018 - Issue 4.

[26] Krasnova, Abramova, et al (2016). Why phubbing is toxic for your relationship: Understanding the role of smartphone jealousy among "Generation Y" users. Research Papers - 109.

[27] Halpern & Katz (2017). Texting's consequences for romantic relationships: A cross-lagged analysis highlights its risks. Computers in Human Behavior - V 71, 06-2017, P 386-394.

[28] Wang, Xie, Wang, Wang y Lei (2017). Peer relationship and adolescent smartphone addiction: The mediating role of self-esteem (…). Journal of Behavioral Addictions Vol 6 – Issue 4.

[29] Guazzini, Duradoni, Capelli, Meringolo (2019). An Explorative Model to Assess Individuals' Phubbing Risk. Future Internet 11(1):21 January 2019.

[30] Kircaburun & Griffiths (2018). Instagram addiction and the Big Five of personality: The mediating role of self-liking. Journal of Behavioral Addictions Volume 7: Issue 1.

[31] Karadag et al. (2015). Determinants of phubbing, which is the sum of many virtual addictions: A structural equation model. Journal of Behavioral Addictions Volume 4: Issue 2.

[32] Casadesus, Fernández, Sequeira, Quintana, Saldana, Ruiz (2012). IPTV Quality assessment system. LANC '12: Proceedings 7th LATAM Networking Conf. Oct 2012 P 52–58.

[33] Blasco, Aznar, Hernández, Ruiz (2011). IPTV as a services distribution channel. Importance of interactivity personalization in the purchasing of news-on-demand packages. Ind. Mngmnt & Data Systems Vol 111 N 9, 2011 pp. 1381-1398.

[34] Mora, Fernández, Ruiz, Cebollero (2021) Cloud QoX 4 EDU & more. VI QQCM QoS y QoE en comunicación multimedia. ISBN: 978-84-09-31124-8.

[35] Greengrass, Evans, Begen (2009). Not All Packets Are Equal, Part 2. The Impact of Packet Loss on Video Quality. IEEE Internet Computing Volume 13, Issue 2, March-April 2009.

[36] Greengrass, Evans, Begen (2009). Not All Packets Are Equal, Part 1. Streaming Video Coding and SLA Requirements. IEEE Internet Computing Volume 13, Issue 1, Jan-Feb 2009.

[37] Klaue, Rathke, Wolisz (2003). EvalVid – A Framework for Video Transmission and Quality Evaluation. TOOLS 2003, LNCS 2794, pp. 255–272, 2003.

[38] Cristobo, Zabala, Ibarrola, Ferro, Liberal (2019). Metodología para la gestión de la QoX basada en el aprendizaje automático. XIV Jornadas de Ing. Telemática (JITEL 2019).

[39] Lobo, López, Mora et al (2008) Distribution Network as comm. system. CIRED Frankfurt. CD 978-1-84919126-5.

[40] Mora, López et al (2008) Smart communications in demand management. CIRED Frankf DOI 10.1049/cp.20090710

[41] Lobo, López, Cabello, Mora et al (2009) How to design a communication network over distribution networks. CIRED Prague. CD:978-1-84919126-5.

[42] Mora, López, Román, Lobo, Carmona, Cabello et al (2009) Demand management communications architecture. CIRED Prague. DOI: 10.1049/cp.2009.0710.

[43] López, Román, Mora, et al (2009) Communications architecture of smart grids to manage the electrical demand. IEEE Int Symp Power Line Comms (ISPLC). Udine, ITALY. Corpus ID: 18821472

[44] Ramírez, Ordiales, Mora et al (2012) Smart Grid - Demand Management as key resource for improvement and social contribution to 2020 strategy. CIGRE Paris. P12-0283/C6_117_2012

[45] Mora, Oyarzábal, Cruz, González, Corera (2012) E-car and economic impact: Enhancing the smart grids. IET Lisbon. ISBN:978-1--84919-628-4.

[46] Oyarzábal, Rodríguez, Cruz, Corera, Mora et al (2013) E-Car and Economic Impact in Smart Grids. CIRED. Stockholm. ISBN:978-1--84919-628-4.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Plataforma modular para la codificación y distribución interactiva de contenidos VR360 basada en campo de visión

Miguel Fernández-Dasi[1], Miguel A. Torres-Font[2], Mario Montagud[1,2] y Miguel Garcia-Pineda[2].

[1]Media & Internet Area
Fundación i2CAT
C/ Gran Capità 2-4 Edifici Nexus I, Barcelona

[2]Departamento de Informática
Universitat de València
Av. De la Universitat, s/n. Burjassot. Valencia

miguel.fernandez@i2cat.net, mitofont@alumni.uv.es, mario.montagud@i2cat.net, miguel.garcia-pineda@uv.es

El consumo de contenidos multimedia ha aumentado mucho en los últimos años. Ello incluye especialmente a los sistemas que ofrecen contenidos inmersivos, como el vídeo VR360, ya que ofrecen experiencias más realistas a los usuarios finales. Pero estos sistemas demandan unos requisitos de red, procesamiento, etc. que suponen desafíos desde el punto de vista técnico siempre y cuando se quiera ofrecer contenidos de alta calidad. Para superar estos retos y limitaciones en este artículo se presenta una plataforma modular extremo-a-extremo para la captura, aplicación de varias estrategias de proyección, trans-codificación, conversión a Dynamic Adaptive Streaming over HTTP (DASH), distribución de baja latencia y consumo interactivo de vídeos VR360, así como para la medida de una gran variedad de métricas Quality of Service (QoS) / Quality of Experience (QoE) de interés. La plataforma constituye un testbed idóneo para investigar sobre ciertos aspectos y procesos esenciales a lo largo de la cadena de distribución, como son la codificación y distribución de los contenidos, teniendo en cuenta potenciales regiones de interés (Regions of Interest (RoI)) en los vídeos y los campos de visión (Field of View (FoV)) de los dispositivos de consumo utilizados.

***Palabras Clave*- VR360, DASH, Region of Interest (RoI), Field of View (FoV).**

I. INTRODUCCIÓN

La producción y consumo de contenidos multimedia se han incrementado progresivamente en los últimos años. Asimismo, la creciente resolución de dichos contenidos, así como su necesidad de distribuirlos en entornos heterogéneos (ej. redes y dispositivos de consumo con capacidades muy variables), no sólo requieren la disponibilidad de anchos de banda y capacidades de procesado suficientes, sino también de técnicas de codificación y distribución adaptativa (ej. basadas en streaming adaptativo HTTP). Estos requisitos son, a su vez, más críticos cuando se refiere a contenidos de Realidad Virtual (RV), tales como el vídeo 360º (en adelante VR360), pues capturan mayor cantidad de información y suelen tener resoluciones altas con el fin de proporcionar un nivel de inmersión satisfactorio.

La comunidad científica sigue muy activa con tal de superar retos y limitaciones existentes para proporcionar de manera más eficiente servicios distribuidos de vídeo VR360 [1, 2]. Este artículo se centra en esta temática, presentando una plataforma extremo-a-extremo modular que posibilita un banco de pruebas de investigación sobre aspectos esenciales como son la captura, procesado, distribución y consumo interactivo de vídeo VR360. En particular, este artículo presenta evidencias preliminares sobre los beneficios que pueden aportar el diseño y adopción de soluciones innovadoras y adaptativas basadas en el campo de visión (*Field of View* (FoV)) y/o regiones de interés (*Region of Interest* (RoI)), en cuanto a el sistema de proyección 360º (ej. *Equirrectangular Projection* (ERP) vs *Cubemap Projection* (CMP)) y a la codificación de vídeo. Estas soluciones se basan en la premisa de que no toda la información capturada en el panorama 360º tiene la misma relevancia, así como que los usuarios sólo pueden visualizar en todo momento una porción del espacio 360º, determinado por el FoV de su dispositivo de consumo, que puede oscilar entre 90º-130º. Además, la plataforma se complementa con una serie de herramientas para el registro de métricas de *Quality of Service* (QoS) indicadoras no sólo del consumo de recursos de cada estrategia y componente bajo análisis, sino también de su impacto sobre la *Quality of Experience* (QoE) percibida.

[1] https://zenodo.org/record/3909281#.YMz-TmgzaiN

II. TRABAJOS RELACIONADOS

En los últimos años la comunidad científica y la industria han dedicado esfuerzos considerables en el área referente a sistemas de vídeo VR360 streaming, proponiendo optimizaciones a lo largo de la cadena extremo-a-extremo [1, 2], incluyendo diversas alternativas en cuanto a estrategias de proyección y codificación basadas en FoV / RoI. A continuación, se destacan brevemente algunas de las contribuciones más relevantes en el área, relacionadas con el trabajo que se presenta. En [3] se presenta una plataforma extremo-a-extremo VR360 que soporta ERP y CMP, así como la medida de métricas, pero no soporta codificación basada en RoI. En [4] se presenta una solución basada en proyección ERP que permite concentrar la calidad en la RoI, pero requiriendo una superposición de tramas. En [5] se propone concentrar la resolución en una determinada RoI cuando se usa CMP, pero utilizando diferentes flujos de vídeo. En [6] se propone una solución para concentrar la calidad dinámicamente en la RoI cuando se utiliza CMP, mediante estimaciones de proyecciones para cada posible dirección/orientación de la cámara virtual. También se han tratado de proponer alternativas más avanzadas que ERP y CMP, como es *Equi-Angular Cubemap* (EAC) [7]. EAC ha sido adoptada por Youtube y proporciona mejoras en cuanto a la uniformidad de densidad de píxeles, aunque requiere una mayor complejidad computacional.

En general, las soluciones existentes se basan en el uso de códecs con restricciones de licencia, requieren el uso de varios codificadores y/o flujos (ej. codificación escalable, multi-nivel, etc.), y presentan problemas de compatibilidad con navegadores web (ej. uso de H.265). En este trabajo se persiguen soluciones más simples, pero efectivas, basadas en el uso de códecs tradicionales como H.264, que son compatibles con los navegadores web, y en procesos de codificación y flujos multimedia únicos. Además, se persiguen soluciones que impliquen modificaciones y cargas mínimas en el reproductor. Ello no sólo permitirá una mayor escalabilidad e interoperabilidad, sino también un mejor rendimiento en sistemas interactivos.

III. PLATAFORMA VR360

En esta sección se presenta la plataforma VR360 desarrollada (ver Fig. 1), con sus bloques / módulos principales: preparación de contenidos (CP), servidor de contenidos (CS), señalización y distribución de contenidos (CD) y consumo de contenidos (CC).

A. Preparación de contenidos – CP

La plataforma puede tomar como entrada videos 360º almacenados o en vivo. En cuanto a la proyección de los vídeos VR360, no sólo se soporta las proyecciones típicas ERP y CMP uniformes (parte superior de Fig.2), sino que se han propuesto soluciones innovadoras para concentrar la resolución en determinadas RoI (parte inferior de Fig.2 para CMP y Fig.3 para ERP). En el caso de CMP, se propone un re-ajuste del cubo para, con una determinada relación de aspecto, concentrar la resolución en una o varias caras. En el caso de ERP, se pueden aplicar dinámicamente diferentes ajustes de codificación a la región exterior al RoI (Fig.3) gracias al desarrollo de un sistema y algoritmo de codificación interactivo y adaptativo, alimentado a partir de patrones de visionado enviados des del reproductor multimedia.

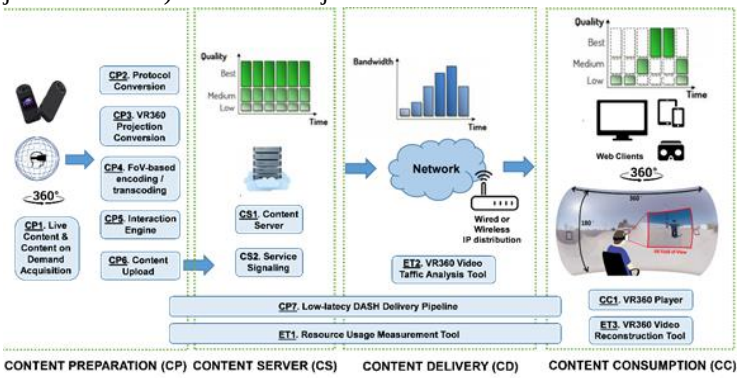


Fig. 1. Diagrama de la plataforma VR360 desarrollada

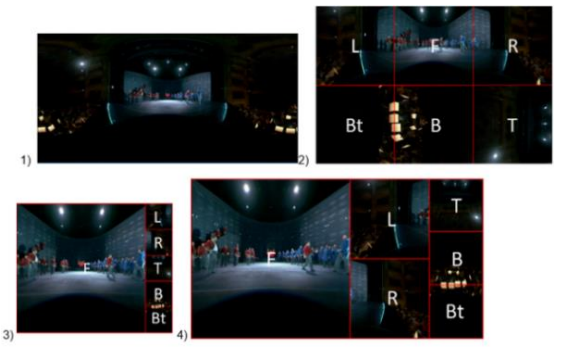


Fig. 2. Proyecciones VR360 consideradas: 1) ERP; 2) CMP 3:2; 3) CMP 6:5 concentrando la resolución en la cara frontal (F); 4) CMP 11:6 concentrando la resolución en el panorama horizontal

Fig. 3. Codificación basada en RoI aplicada a proyección ERP

A continuación, el flujo de vídeo se codifica y se convierte a Dynamic Adaptive Streaming over HTTP (DASH) [7, 8], con los ajustes deseados, e integrando sus recientes funcionalidades para su distribución en baja latencia. En el caso de recepción de vídeos en vivo, este bloque también soporta la conversión entre protocolos origen y destino, así como la trans-codificación del flujo. Finalmente, el vídeo resultante se transfiere a un servidor (web) de contenidos (ej. Apache).

Las soluciones propuestas no están restringidas al uso de códecs avanzados que impliquen problemas de licencias y/o compatibilidad en ciertas plataformas y/o navegadores web, ni requieren el uso de varios flujos de transmisión, aportando así ventajas significativas para su adopción.

B. Señalización y Distribución de contenidos – CS y CD

Además de los contenidos, el servidor web aloja un fichero para la indexación de los mismos, incluyendo información relevante sobre los mismos (ej. título, duración, tipo de proyección…).

C. Consumo de contenidos – CC

En cuanto al consumo de contenidos, se ha adoptado un reproductor web *open-source* [7] y extendido con tal de: 1) procesar las proyecciones basadas en RoI propuestas (a partir de la información proporcionada por el fichero de indexación); 2) detectar y reportar periódicamente al sistema de codificación los patrones de visionado (centro del FoV), cuando se utiliza la proyección ERP basada en RoI. Una ventaja fundamental de las soluciones basadas en RoI propuestas en este trabajo es que no requieren modificaciones en el reproductor en caso de adoptar ajustes de RoI pre-establecidos (ej. caras del cubo) o mínimas en el caso de requerir ajustes de RoI dinámicos, simplemente teniendo que reportar el centro del FoV a través de una conexión vía sockets con el sistema de codificación del bloque de preparación de contenidos.

IV. *Herramientas de Medida de Métricas*

La plataforma se complementa con una serie de herramientas para la medida de métricas QoS/QoE, con tal de determinar los beneficios de las soluciones propuestas.

A. Medida de Consumo de Recursos

Herramienta software *open-source* [9] que permite medir el uso de CPU (%), GPU (%) y memoria RAM (MB) para procesos de Windows, haciendo uso de llamadas a la *Powershell*. Puede ser aplicada a cada componente de la plataforma, aunque resulta de especial interés para los módulos de (trans-)codificación y consumo de contenidos.

B. Análisis de tráfico DASH

Herramienta software *open-source* [9] para capturar y analizar el tráfico DASH, así como las peticiones HTTP asociadas, para una sesión de streaming determinada, haciendo uso de *tshark* (versión de consola de *WireShark*). Esta herramienta no sólo captura el fichero *Media Presentation Description* (MPD) asociado al vídeo DASH seleccionado, sino también los parámetros asociados a cada segmento DASH solicitado/descargado (URL, códec, resolución, *bitrate*, *fps*…), así como su tamaño, número de segmentos asociados y retardo de ida y vuelta .

C. Medida de Patrones de Visionado VR360

A partir de las medidas periódicas del centro del FoV en el reproductor, se ha desarrollado un script para representar el mapa de calor de las regiones / patrones de visionado en un plano 2D, en forma rectangular (para ERP) o de T tumbada (para CMP).

V. Tests y resultados Preliminares

En esta sección se presentan brevemente algunos de los tests realizados, así como los resultados preliminares obtenidos, cuando: 1) se compara la proyección ERP con CMP, y sus variantes; y 2) se analiza el potencial impacto de aplicar codificación basada en RoI para ERP. La capacidad de la plataforma para registrar métricas QoS, tales como la calidad DASH seleccionada, el *throughput* estimado o el nivel de ocupación del buffer de reproducción, se demostró en [3].

A. Comparativa entre ERP y (variantes de) CMP

En primer lugar, se codificaron 3 vídeos DASH con los mismos niveles de calidad (resoluciones y bitrates) equivalentes, en función de sus relaciones de aspecto, para comparar tanto a nivel QoS como QoE las siguientes condiciones de test (TC): *TC1*) ERP con CMP; *TC2*) CMP 3:2 uniforme con CMP 6:5; y *TC3*) CMP 3:2 uniforme con CMP 11:6. En cuanto a eficiencia de compresión, los resultados confirman que se puede conseguir un ahorro de tamaño de vídeo y, por tanto, de capacidad en servidor y de ancho de banda, entre 20-25% cuando se usa CMP con respecto al uso de ERP, ambos en sus variantes uniformes. Estos resultados están en línea con los obtenidos en el estado del arte [1, 2]. A su vez, las variantes de CMP basadas en RoI propuestas introducen una ganancia con respecto a CMP uniforme entre el 10-15% de tamaño.

Además, se realizaron pruebas subjetivas con usuarios (*N=24* participantes) comparando las tres condiciones de test, alternando el orden de presentación de los vídeos a consumir. En dichas pruebas, se presentaron cuestionarios sobre percepción QoE (utilizando medidas *Mean Opinion Score*) e inmersión (*Igroup Presence Questionnaire (IPQ)* [10]). De manera interesante, los resultados indican que a pesar de la reducción de tamaño (y por tanto ahorro de recursos y costes) el uso de CMP no tiene un impacto negativo sobre los niveles de QoE e inmersión percibidos, así como que el uso de las variantes CMP basadas en RoI pueden incluso aportar mejoras, en función del vídeo.

[1] https://zenodo.org/record/3909281#.YMz-TmgzaiN

Finalmente, la Fig. 4 representa el mapa de calor de los patrones de visionado agregados de los 24 usuarios para el mismo vídeo, codificado en proyección ERP y CMP. Aunque se trate de un vídeo con escenas relevantes a lo largo del entorno 360º, el visionado se concentra en regiones específicas, vislumbrando los potenciales beneficios del procesado y distribución basado en RoI.

B. Uso de codificación basada en RoI para ERP

En segundo lugar, se realizaron unas iteraciones de tests para analizar preliminarmente el impacto de aplicar en un vídeo determinado (resolución 1280x720, duración 1'30'') diferentes valores de Constant rate factor (CRF) para las regiones exteriores a una RoI determinada (posición fija y tamaño variable en cada iteración de test: 1/4, 1/3 y 1/2 de la resolución del vídeo), en cuanto al tamaño del vídeo y a métricas objetivas de calidad, como Structural Similarity Index (SSIM). Los resultados se pueden observar en la Fig. 5.

Por un lado, en términos de tamaño, el uso de la RoI puede resultar en una reducción de hasta un 500% manteniendo índices de calidad prometedores. En la Fig. 5 se puede observar que a medida que se disminuye el tamaño de RoI, también disminuye el tamaño del vídeo, ya que existe una mayor región de vídeo con unos niveles de calidad inferiores. El tamaño de la RoI dependerá del tipo del dispositivo de consumo utilizado, del tipo de vídeo, e incluso podría ajustarse en base a los patrones de visionado. Por otro lado, las gráficas de la Fig. 5 muestran que para rangos de CRF hasta alrededor de 27, el valor de SSIM obtenido se acerca a la unidad, consiguiendo un ahorro de tamaño en torno al 200%. Estos resultados reflejan claramente el potencial beneficio de aplicar codificación basada en la RoI, especialmente en el caso de vídeos VR360 en los que las regiones exteriores a la RoI se pueden ajustar para que se encuadren fuera del FoV.

VI. Conclusiones y trabajos futuros

Este artículo ha presentado una plataforma modular extremo-a-extremo para la captura, aplicación de varias estrategias de proyección, trans-codificación, conversión a DASH, distribución y consumo interactivo de vídeos VR360, así como para la medida de métricas QoS / QoE. La plataforma constituye un *testbed* idóneo para investigar sobre ciertos aspectos y procesos esenciales, como son la codificación y distribución de los contenidos, teniendo en cuenta potenciales RoI y FoV. Se han presentado evidencias preliminares sobre los potenciales beneficios que aportan este tipo de estrategias.

Como trabajo futuro, se va a investigar sobre servicios de streaming VR360 aplicando este tipo de técnicas y estrategias basadas en RoI / FoV para una gran variedad de vídeos y escenarios, así como se evaluarán sus prestaciones en cuanto a latencia y escalabilidad. Finalmente, se especificarán estrategias de modelado QoE para servicios de VR360 streaming basados en DASH [8].

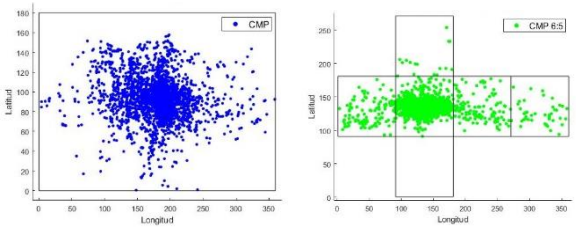


Fig. 4. Mapa de calor de proyección ERP (izq) y CMP (der).

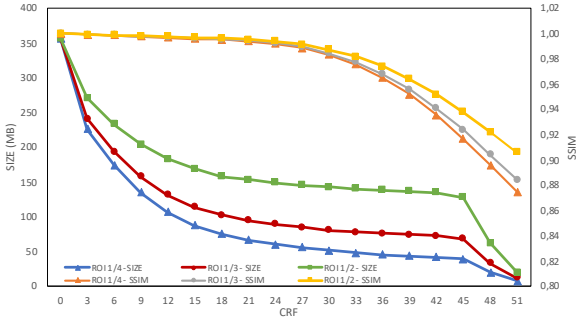


Fig. 5. Análisis de tamaños en MB frente al valor CRF utilizado en codificación de video.

Agradecimientos

Este trabajo ha sido financiado parcialmente por la Generalitat Valenciana en el marco del proyecto GV/2020/052, por la Comisión Europea en el marco del proyecto EU H2020 Respond-A (ID 883371), por la red temática Everest (RED2018-102383-T) y por la UV (UV-INV-AE-1564749). Asimismo, el trabajo de Miguel Fernández ha sido financiado por la Secretaria d'Universitats i Recerca de la Generalitat de Catalunya i del Fons Social Europeu (Personal Novel 2021 FI_B_01041) y el de Mario Montagud por una Beca postdoc JdC-Incorporación (MICINN, IJCI-2017-34611).

Referencias

[1] C.L. Fan, W.C. Lo, Y.T. Pai, C.H. Hsu, "Survey on 360° Video Streaming: Acquisition, Transmission, and Display", ACM Comput. Surv. 52, 4, Article 71, Sept. 2019, 36 pages

[2] M. Xu, C. Li, S. Zhang, P.L. Callet, "State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression", IEEE JSAC, vol. 14, no. 1, pp. 5-26, Jan. 2020

[3] M. Montagud, E. Meyerson, I. Fraile, S. Fernández, "Modular Testbed for KPI Monitoring in Omnidirectional Video Streaming Scenarios", MMEDIA 2019, València (Spain), March 2019

[4] A. Xu, X. Chen, Y. Liu and Y. Wang, "A Flexible Viewport-Adaptive Processing Mechanism for Real-Time VR Video Transmission", IEEE ICMEW, 2019.

[5] D. Gómez, J. A. Núñez, I. Fraile, M. Montagud, S. Fernández, "TiCMP: A lightweight and efficient Tiled Cubemap projection strategy for Immersive Videos in Web-based players", ACM NOSSDAV'18, Amsterdam, June 2018

[6] P. Yu. Timokhin, M.V. Mikhaylyuk, E.M. Vozhegov, "Efficient methods and algorithms to synthesize 360-degree video based on cubemap projection of virtual environment", ISP RAS, 32:4, 2020

[7] J. Lin, et al., "Efficient Projection and Coding Tools for 360° Video", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 84-97, March 2019

[8] M. Garcia-Pineda, J. Segura-García; S. Felici-Castell; A. Soriano-Asensi M. Montagud, "Modelando la QoE en entornos DASH", JITEL 2019, Zaragoza (Spain), October 2019

[9] M. Montagud, J. Antonio De Rus, R. Fayos-Jordán, M. Garcia-Pineda J. Segura-Garcia, "Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics", ACM MMSYS 2020, Istanbul (Turkey), June 2020

[10] Igroup Presence Questionnaire (IPQ), http://www.igroup.org/pq/ipq/index.php Acceso en 30/09/2021

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Holo-conferencias 3D multi-usuario: hacia una nueva generación de reuniones virtuales

Sergi Fernández[1], Mario Montagud[1], Gianluca Cernigliaro[1], Marc Martos[1], David Rincón[2]

[1]Media & Internet Area
Fundación i2CAT
C/ Gran Capità 2-4 Edifici Nexus I, Barcelona

[2]Departament d'Enginyeria Telemàtica
Universitat Politècnica de Catalunya (UPC)
Edifici C4C, Castelldefels (Barcelona)

{sergi.fernández; mario.montagud; gianluca.cernigliaro; marc.martos}@i2cat.net; david.rincon@upc.edu

Los sistemas de videoconferencia multi-usuario han ganado mucha relevancia en la sociedad. Aunque estos hayan evolucionado considerablemente en cuanto a escalabilidad, interoperabilidad y funcionalidades de interacción, todavía presentan limitaciones importantes en cuanto a realismo, calidad de interacción y confort, debido principalmente al uso de representaciones de bustos 2D encuadrados en una matriz. Este artículo presenta una plataforma extremo-a-extremo que posibilita una nueva era para las reuniones virtuales mediante la integración en tiempo real de representación volumétricas holográficas de los usuarios en entornos 3D compartidos, mediante tecnologías compatibles a los estándares actuales y utilizando equipamiento de bajo coste. El artículo además reporta sobre resultados obtenidos, tanto a nivel de rendimiento y consumo de recursos para escenarios típicos, como en lo que se respecta a la experiencia de usuario en sesiones de holo-conferencia interactivas de cuatro participantes.

***Palabras Clave*- Holo-Conferencias, Hologramas, QoS, QoE, Realidad Virtual Social, Video Volumétrico**

I. INTRODUCCIÓN

Los avances tecnológicos en las últimas dos décadas han permitido la proliferación de herramientas de comunicación, en tiempo real y multiusuario, como son los sistemas de videoconferencia que permiten reuniones remotas e incluso el tele-trabajo, a un coste mínimo y suponiendo una alternativa real a la presencialidad.

En el campo de la investigación se han realizado esfuerzos considerables con tal de optimizar los sistemas de videoconferencia (ej. reducción de latencias, compresión de video y audio, escalabilidad para sesiones multiusuario…) [1], así como para determinar el impacto sobre la calidad de la experiencia (QoE) que provocan ciertos umbrales en cuanto a parámetros de calidad de servicio (QoS) [2], que en gran medida vienen a determinar el coste del servicio. A su vez, estos sistemas de videoconferencia se han ido complementando con funcionalidades interactivas que mejoran y enriquecen la experiencia compartida [3], como serían la compartición de ficheros y/o pantalla, el ajuste del fondo y composición de la representación 2D de los participantes, o incluso sincronizar la reproducción de contenido compartido.

Más recientemente, la madurez de los sistemas de Realidad Virtual (RV) y su convergencia con los sistemas de audio-videoconferencia están permitiendo el desarrollo de una nueva generación de plataformas de Social VR (ej. Mozilla Hubs, Altspace, Facebook Horizon, Spatial, Glue…) que permiten la interacción y la comunicación entre múltiples usuarios remotos inmersos en entornos virtuales compartidos y, aunque no es condición *sine qua non*, a través de cascos de Realidad Virtual o Aumentada (RV o RA). Gracias a estas plataformas, el lienzo donde se plasma la comunicación ya no está necesariamente restringido a una pantalla rectangular donde se representan múltiples usuarios y contenidos, sino que el entorno (3D) digital con libertad de exploración, o incluso el entorno real en el caso de la RA, se convierte en dicho lienzo, con las nuevas oportunidades y retos que ello supone. Ejemplos claros de estos retos son el método de representación de los usuarios y cómo se interactúa con otros usuarios y el entorno.

En 2020, con la llegada de la pandemia mundial, los sistemas de videoconferencia se han convertido en una herramienta fundamental, extendiendo su uso ampliamente y en ámbitos que trascienden el profesional, permitiendo la socialización en un contexto de distanciamiento social obligatorio. Del mismo modo, las plataformas de Social VR se han ido popularizando y su uso se ha visto acelerado por la necesidad de una mayor y mejor interacción que la que ofrecen hoy en día los sistemas de videoconferencia 2D. A pesar del potencial que estas plataformas pueden ofrecer para una comunicación más natural y realista, hay pocos estudios al respecto, y los que hay se enfocan principalmente en aspectos tecnológicos como la trasmisión de contenido volumétrico [4], el consumo de contenidos compartidos en

entornos virtuales [5], o el nivel de identificación con avatares virtuales (Sección II), pero dejando de lado aspectos clave como son la representación de los usuarios y cómo esta afecta al proceso de comunicación entre ellos, cuando este proceso es, precisamente, el principal objetivo en experiencias de Social VR.

Este articulo presenta una evolución de una plataforma Social VR [4, 5] de bajo coste en el cual los usuarios son capturados por una o múltiples cámaras con sensores de profundidad, esto es RGB-D (ej. Azure Kinect), e integrados en tiempo real y en un formato fotorrealista y volumétrico (nubes de puntos 3D) en entornos virtuales compartidos. Tras presentar los componentes tecnológicos que componen la plataforma, se evalúa su uso cuando se utilizan representaciones de usuarios capturados por una simple cámara RGB-D tanto a nivel QoS como QoE en sesiones interactiva en grupos de cuatro participantes.

En la Figura 1 se representa un ejemplo de reunión virtual por holo-conferencia utilizando la plataforma Social VR desarrollada, así como del equipamiento utilizado por usuarios en los tests realizados. Los resultados obtenidos sirven para determinar los requisitos y costes de implantación de este tipo de servicios, así como demostrar su potencial impacto debido al alto interés que suscitan y a los prometedores niveles de (co-)presencia y calidad de interacción que este nuevo *medium* de comunicación es capaz de proporcionar.

II. Estado del Arte

Las soluciones de Social VR pretenden proporcionar experiencias similares a las reuniones presenciales, tratando de maximizar la sensación de presencia, co-presencia, así como la calidad de la interacción y la plausibilidad del conjunto. En [6] se analizan los factores que tienen impacto en la sensación de identificación con un avatar (*embodiment*). En [7] se demuestra como una sincronización entre acciones del propio cuerpo y su representación mediante un avatar incrementa la sensación de presencia en un entorno virtual. En [8] se demuestra que la representación mediante avatares realistas mejora la sensación de *embodiment* y presencia, quedando confirmado en otros estudios posteriores (ej. [9, 10]). El estudio en [11] aporta evidencias sobre un incremento de presencia, emoción y reconocimiento de los usuarios cuando se usan avatares a partir de reconstrucciones 3D realistas obtenidas en tiempo real, en comparación al uso de avatares sintéticos animados. Asimismo, en [12] se aportan evidencias preliminares sobre los potenciales beneficios que pueden aportar las capturas volumétricas mediante sensores RGB-D y cascos de RV en el ámbito de las experiencias de consumo multimedia compartidas. En [13] se presenta una plataforma Social VR en la que los usuarios se representan en formato vídeo RGB-D mediante el uso de cámaras únicas, y se integran en un entorno estático formado por una imagen 360º. En [14] se desarrolla un cuestionario para experiencias Social VR y se utiliza para comparar experiencias de visionado de fotos compartidas en grupos de 2 usuarios, utilizando tres condiciones de test: a) escenario físico real; b) uso de Skype; y c) plataforma Social VR desarrollada en [13]. Se concluye que el uso de Social VR con representaciones realistas de los usuarios mejora la experiencia de usuario (presencia, co-presencia y calidad de interacción) en comparación al uso de Skype, así como proporciona experiencias comparables a entornos físicos. En [5] se confirman estos beneficios en un escenario de visionado de vídeo compartido, también en grupos de 2 personas, cuando se utiliza Facebook Spaces (plataforma Social VR en la que los usuarios se representan como avatares) como condición de test de comparación. Asimismo, en [5] se presenta una versión evolucionada de la plataforma en [13], incorporando representaciones volumétricas de los usuarios, así como entornos virtuales compartidos en 3D, tras haber realizado un análisis del estado del arte y haber comprobado que existen únicamente un par de plataformas con dichas características, pero requieren equipamiento complejo y caro, y no tienen un soporte multiusuario claro. Finalmente, en dicho estudio se demuestra los potenciales beneficios que esta tecnología de holo-portación puede aportar en una gran variedad de casos de uso. Finalmente, en [4] se presenta una versión evolucionada de la plataforma en [5] con soporte para más de 2 usuarios, y para representaciones volumétricas de los usuarios mediante nubes de puntos (*Point Clouds*). Esta versión de la plataforma Social VR se adopta en este trabajo para investigar el potencial y nivel de madurez de esta tecnología puntera y de bajo coste para posibilitar servicios de holo-portación (esto es, tele-transportación holográfica) multiusuario en tiempo-real, no centrándose únicamente en entornos de consumo de contenidos compartidos como se ha hecho en trabajos anteriores, sino poniendo el foco en la calidad de los usuarios y de la comunicación/interacción entre los mismos.

Fig. 1. Escenarios de holo-conferencia desarrollados y evaluados en el artículo: a) captura de pantalla de sala de reuniones virtual con cuatro participantes; b) y c) equipamiento por cada usuario, con sistema de captura con cámara RGB-D única y casco RV.

III. Plataforma Social VR

Esta sección presenta la plataforma desarrollada a partir de [4] y [5]. A diferencia de dichos trabajos previos, la plataforma que se presenta soporta más de dos usuarios capturados en tiempo real e insertados en el mismo entorno virtual compartido, amplía el abanico de tipologías de representaciones de usuarios soportadas (aunque este artículo se centre en nubes de puntos capturadas por sensores RGB-D únicos), permite la ingesta y reproducción de contenidos pre-grabados y en vídeo, y soporta consumo en modo pantalla 2D y en cascos RV. Además, incorpora una gestión de sesiones concurrentes y una gestión de eventos que permite la interactividad con el entorno más allá de la visualización de contenidos. En la Figura 2 se muestra un diagrama aproximado sobre la arquitectura del sistema. Las siguientes subsecciones describen más en detalle sus componentes principales.

A. *Captura y transmisión de nubes de puntos*

El cuerpo de los usuarios se captura mediante 1 o varias cámaras RGB-D (ej. Kinect Azure), basándose en el sistema presentado en [4], que convierte las diferentes imágenes de color y profundidad capturadas por el sensor RGB-D en una nube de puntos 3D (fusionada, en el caso de utilizar sistemas de captura de varios sensores). Por un lado, la nube de puntos se renderiza en el reproductor (Sección III.C) que se ejecuta en el mismo entorno local que el sistema de captura para posibilitar la representación del usuario local y que, por tanto, se vea su propio cuerpo (*self-representation*). Por otro lado, la nube de puntos se comprime usando el códec propuesto en [15] y se transmite vía un Orquestador (Sección III.B), utilizando bien *Dynamic Adaptive Streaming over HTTP* (DASH) o una comunicación basada en sockets mediante *socket.io*. El códec adoptado [15] permite la compresión de nubes de puntos usando la ocupación de los nodos en *octrees* para representar la geometría y proyectando el color de cada nodo en un *grid* 2D. A pesar de que el códec permite explotar las inter-dependencias temporales entre tramas, en este trabajo se utilizan únicamente tramas *intra* para no comprometer la latencia del sistema.

B. *Orquestador*

Los componentes de orquestación son comúnmente usados en sistemas de videoconferencia para gestionar sesiones y flujos de datos [16]. El orquestador desarrollado maneja la información relativa a cada participante de la sesión, así como de los flujos de datos que intervienen en la misma (audio, video 2D / volumétrico, eventos, etc.), actuando como un reflector. Asimismo, se encarga de notificar sobre cambios en el estado de la sesión (ej. cambios de posición). Finalmente, el orquestador se encarga de gestionar la distribución interactiva de los contenidos asociados a cada sesión (ej. un video a presentarse en una pantalla virtual en el entorno 3D).

C. *Reproductor Multimedia*

El reproductor (*player*) de la experiencia RV se ha desarrollado en Unity (ejecutable Windows) y es el responsable de la interacción con el Orquestador, instanciar los sistemas de captura y de transmisión/recepción, así como de representar todos los flujos que componen la experiencia adecuadamente, incluyendo el entorno virtual compartido, que puede estar alojado en el mismo *player* o bien descargarse de un servidor de contenidos al inicio de la sesión.

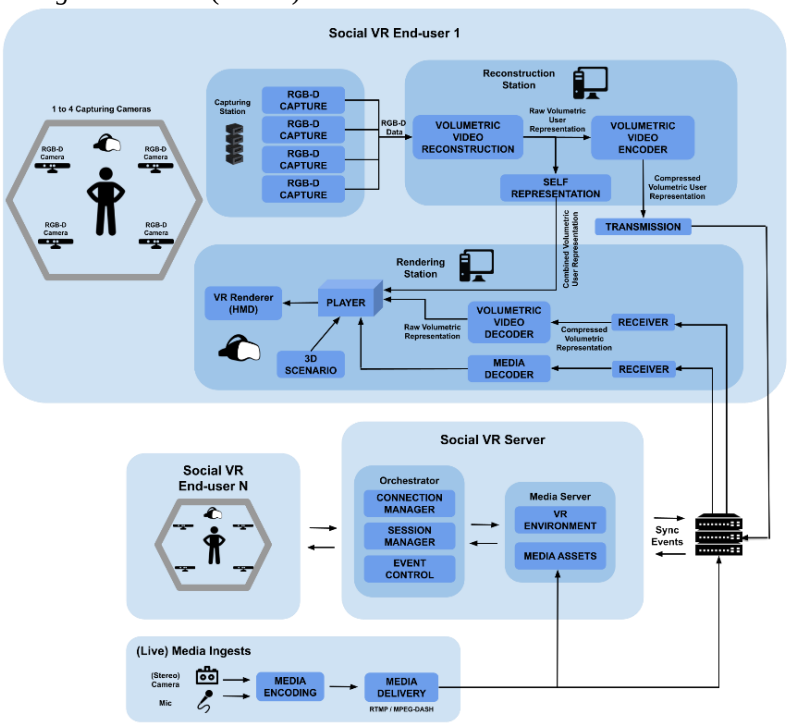


Fig. 2. Arquitectura y diagrama de flujo de alto nivel de la plataforma Social VR desarrollada

IV. Evaluación y Resultados

Esta sección reporta sobre resultados de evaluaciones objetivas (QoS) y subjetivas (QoE) realizadas, con el objetivo de determinar los requisitos y rendimiento de la plataforma, así como la receptividad, calidad percibida e interés en estos escenarios, respectivamente.

Las pruebas reportan sobre experimentos con grupos de 4 usuarios, capturados por cámaras RGB-D únicas, a 15 fps y aproximadamente 50000 puntos por trama, y utilizando *socket.io* para su distribución. En cuanto a los parámetros QoS, se midieron en PCs con los siguientes recursos: CPU Intel(R) Core(TM) i7-10750H @ 2.60GHz, 16GB de RAM y una GPU NVIDIA GeForce RTX 2070.

A. Evaluación Objetiva

En primer lugar, se midió el uso recursos computacionales en uno de los PC clientes (a temperatura ambiente), utilizando un herramienta desarrollada en [17]: CPU: 22.86 %; GPU: 34.53 %; y RAM: 630.11 MB. Para el número fijado de puntos por trama y tramas por segundo por cada usuario, se estima que un PC sin gráfica dedicada suportaría hasta dos usuarios en un escenario simple. En segundo lugar, se midió el consumo de ancho de banda correspondiente a los flujos de nubes puntos. Para los parámetros fijados, cada flujo consumió alrededor de 6.2 Mbps (stdv=1.1 Mbps). En tercer lugar, se midió el retardo extremo-a-extremo (esto es, desde captura hasta renderizado) para cada flujo de nubes puntos, en un escenario en el Orquestador está desplegado en una ciudad situada a 40Km de distancia de los PCs de los clientes finales. Para los parámetros fijados, el retardo para cada flujo fue de 211.22 ms (stdv=10.3 ms). Estos resultados se refieren al promedio para 5 sesiones de unos 8 minutos.

B. Evaluación Subjetiva

Se realizaron tests en grupos de 4 usuarios (*N=32* participantes, edad entre 18-40 años, 17 mujeres), integrándolos en una sala de reuniones virtual, equi-espaciados alrededor de una mesa redonda (Figura 1). Para estimular la interacción, se les instruyó para que realizasen tareas gamificadas, como juegos de adivinanzas (oficios, ciudades, películas…) utilizando gestos no-verbales. La duración de cada sesión fue de 8-10 minutos.

Se adoptó la metodología de evaluación propuesta en [14, 5], incluyendo cuestionarios sobre (co-)presencia, calidad de interacción, así como sobre usabilidad, mareo, cansancio y dificultad. También se realizaron entrevistas con los usuarios. Por motivos de extensión, este artículo se centra en los resultados obtenidos para cuestionarios adicionales diseñados específicamente para capturar la percepción y opinión de los usuarios en cuanto a la calidad audiovisual, comparación con experiencias reales, así como a potencial e impacto de la tecnología que se presenta. Las Tablas I-III reportan los resultados obtenidos, que son muy satisfactorios y prometedores.

V. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo ha presentado una versión evolucionada de una plataforma Social VR y evaluado su potencial para / en escenarios de holo-conferencia o reuniones virtuales multiusuario. A pesar de margen de mejora (ej. en cuanto a la calidad de la reconstrucción gráfica, la falta de un volumen completo en sistemas de captura de cámara RGB-D única, escalabilidad…), los usuarios se han mostrado muy satisfechos con la experiencia, reportando alto niveles de (co-)presencia y calidad de interacción, y sin experimentar mareos ni cansancio. Además, el sistema se comporta de forma satisfactoria y robusta, evidenciando madurez tecnológica.

En cuanto a la experiencia de usuario, en futuros trabajos cabe esperar una comparación con condiciones *baseline* (reunión física, solución de videoconferencia tradicional u otras plataformas de Social VR). Aun así, esta es una primera aproximación que demuestra el potencial de este *medium* es escenarios multi-usuario donde el foco no está tanto en el entorno o en un consumo compartido, sino en los usuarios mismos y en la interacción entre los mismos. A pesar de que en el presente trabajo se ha utilizado un sistema de captura sencillo basado en una sola cara RGB-D frontal, los resultados demuestran que esta configuración puede ser suficiente para entornos en los que los usuarios no se mueven libremente por el entorno, incluso cuando estos están localizados en vistas laterales.

Como trabajo futuro, también se plantea la mejora de las prestaciones del sistema, a varios niveles. Primero, se persigue incrementar la calidad de la representación visual de los usuarios. Segundo, se persigue incrementar la escalabilidad del sistema. Tercero, se pretende habilitar sesiones compartidas con usuarios utilizando diferentes formatos de representación y tipos de dispositivos..

Tabla I
Calidad Audiovisual (1=muy mal; 5=Excelente)

Pregunta / Puntuación	1	2	3	4	5
The visual quality of the virtual scenario	-	-	1	22	9
The visual quality of my representation	-	2	13	15	2
The visual quality of the representation of the user(s) next to me	-	4	16	12	-
The visual quality of the representation of the user(s) in front of to me	-	-	7	21	4
The audio quality from the user(s)	-	-	-	20	12

Tabla II
Comparación a Escenario Real (1=Mucho Peor; 3=Igual; 5=Mucho Mejor)

Pregunta / Puntuación	1	2	3	4	5
The overall virtual experience with one in real life	-	2	19	3	1
The visual representation of users in the virtual experience compared to a real scenario	2	19	11	-	-
The audio quality in the virtual experience compared to the one in a real scenario	-	5	20	7	-
The naturalness of the gestures in the virtual scenario, compared to a real scenario	-	8	22	2	-
The overall communication quality in the virtual scenario, compared to a real scenario	-	5	19	7	1

Tabla III
Potencial e Impacto (1=Totalmente en Desacuerdo; 5=totalmente de Acuerdo)

Pregunta / Puntuación	1	2	3	4	5
This system is effective to hold virtual meetings	-	-	-	6	26
The quality of the users' representation is enough to enable effective and comprehensive interactions and collaborations	-	1	1	21	9
I would use a system like this one for meetings and collaborative tasks in virtual scenarios	-	-	-	10	22
These kind of systems can contribute to a more sustainable environment	-	-	5	12	15

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por el programa H2020 de la Unión Europea, en el marco del proyecto VR-Together (ID 762111) y por ACCIÓ (RIS3CAT, Generalitat de Catalunya), en el marco del proyecto VIVIM (Ref. COMRDI18-1-0008). Asimismo, el trabajo de Mario Montagud ha sido financiado por una Beca JdC-Incorporación (MICINN, IJCI-2017-34611).

REFERENCIAS

[1] S. Firestone, T. Ramalingam, S. Fry, "Voice and Video Conferencing Fundamentals", Cisco Press, 397 pages, March 2007, ISBN-10: 1-58705-268-7

[2] M. Schmitt, J. Redi, D.C.A. Bulterman, P. Cesar, "Towards individual QoE for multi-party video conferencing", IEEE Transactions on Multimedia (TMM), 20(7):1781-1795, 2018

[3] F. Boronat, M. Montagud, P. Salvador, J. Pastor, "Wersync: A web platform for synchronized social viewing enabling interaction and collaboration", Journal of Network and Computer Applications, Volume 175, February 2021.

[4] J. Jansen, S. Subramanyam , R. Bouqueau, G. Cernigliaro, M. Martos, F. Pérez, P. Cesar, "A Pipeline for Multiparty Volumetric Video Conferencing: Transmission of Point Clouds over Low Latency DASH", ACM MMSys 2020, Istanbul (Turkey), June 2020

[5] M. Montagud, J. Li, G. Cernigliario, A. El Ali, S. Fernandez, P. Cesar, "Towards SocialVR: Evaluating a Novel Technology for Watching Videos Together", ArXiv (under review in Virtual Reality (Springer)), arXiv:2104.05060, April 2021

[6] K. Kilteni, R. Groten, M. Slater, "The sense of embodiment in virtual reality", Presence: Teleoperators and Virtual Environments, 21, 4 (2012), 373–387.

[7] P. Heidicker, E. Langbehn, F. Steinicke, "Influence of avatar appearance on presence in Social VR", IEEE 3DUI 2017, 233–234

[8] D. Roth, et al, "Avatar realism and social interaction quality in virtual reality", IEEE VR 2016, 277–278

[9] H. J. Smith, M. Neff, "Communication behavior in embodied Virtual Reality", ACM CHI 2018, 289

[10] T. Waltemate, et al., "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response", IEEE transactions on visualization and computer graphics, 24(4), 1643–1652, 2018

[11] R. Mekuria, P. Cesar, I. Doumanis, A. Frisiello, "Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room", Proc. SPIE 9599, Applications of Digital Image Processing XXXVIII, Vol. 95991, September 2015.

[12] M. McGill, J. H. Williamson, S. Brewster, "Examining the role of smart TVs and VR HMDs in synchronous at-a-distance media consumption", ACM TOCHI, 23, 5, 33, 2016

[13] S. Gunkel, M. Prins, H. Stokking, O. Niamut, "Social VR Platform: Building 360-degree Shared VR Spaces", ACM TVX 2017, Hilversum (The Netherlands), June 2017

[14] J. Li, et al., "Measuring and Understanding Photo Sharing Experiences in Social Virtual Reality", ACM CHI 2019, Glasgow (UK), May 2019

[15] R. Mekuria, K. Blom, P. Cesar, "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video", IEEE Transactions on Circuits and Systems for Video Technology", 27(4), 828–842, 2017

[16] W. Weiss, R. Kaiser, M. Falelakis, "Orchestration for Group Videoconferencing: An Interactive Demonstrator", ACM ICMI '14, Istanbul (Turkey), 2014

[17] M. Montagud, J. Antonio De Rus, R. Fayos-Jordán, M. Garcia-Pineda J. Segura-Garcia, "Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics", ACM MMSYS 2020, Istanbul (Turkey), June 2020

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Herramientas de telemedicina para autocuidado de pacientes y para ayuda a cuidadores. Prototipo para Android

César Fernández[1], María Asunción Vicente[1], Mercedes Guilabert[2], Irene Carrillo[2], Amaya Vara[3], Aurora Mula[3], Kamila Cheikh[3] & José Joaquín Mira[2,3]

[1]Área de Ingeniería Telemática, Escuela Politécnica Superior de Elche, Universidad Miguel Hernández (UMH)

[2]Departamento de Psicologia de la Salud, Universidad Miguel Hernández (UMH)

[3]Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana (FISABIO)

[1,2] (UMH), Avenida de la Universidad s/n Elche (03202) Alicante, España
[3] (FISABIO), Avda. de Catalunya, 21 / (46020) Valencia, España

c.fernandez@umh.es, suni@umh.es, mguilabert@umh.es, icarrillo@umh.es, avara@umh.es, amula@umh.es, kamilacheikh5@gmail.com, jose.mira@umh.es

Se presenta un prototipo de telemedicina centrado en el autocuidado de pacientes cardiometabólicos y la ayuda a sus cuidadores. El sistema completo comprende servidor, base de datos, aplicación para dispositivos Android y acceso a información de dispositivos *wearables*. La aplicación se estructura en dos zonas, una de ellas accesible al paciente y la otra accesible a sus cuidadores. Entre las funcionalidades para pacientes se encuentran la gestión de medicación y alarmas, el escaneo de medicamentos, el control de la actividad física y el control de las constantes vitales. Entre las funcionalidades para cuidadores se encuentran un sistema de intercambio de información para cuidadores que trabajan por turnos, seguimiento GPS para pacientes que lo requieran, y herramientas de información y ayuda. El prototipo es completamente funcional, a la espera de financiación para su comercialización.

***Palabras Clave*- Telemedicina, m-health, apps de salud, Android.**

I. Introducción

Las capacidades de los terminales móviles actuales, así como de los diferentes dispositivos hardware *wearables*, permiten realizar un seguimiento preciso de la medicación, la actividad y las constantes vitales de los usuarios. El reciente estudio presentado en [1] analiza la aplicabilidad de estas tecnologías para cuidar nuestra salud; mientras que un informe publicado en [2] analiza el compromiso de los usuarios con las aplicaciones de salud. También hay análisis específicos sobre el efecto de diferentes aplicaciones móviles en el control de la medicación [3].

No obstante, estas aplicaciones están pensadas exclusivamente para el autocuidado y no se adaptan correctamente a un gran número de pacientes que, además de cuidarse a sí mismos en muchos aspectos, requieren de cuidadores externos (normalmente familiares) para la ayuda con ciertas tareas; así como a pacientes dependientes que requieren ayuda prácticamente continua.

El objetivo del presente trabajo ha sido el diseño y desarrollo de un prototipo de aplicación de Telemedicina pensada tanto para el autocuidado como para la ayuda a los cuidadores, y que resulte útil para diversos tipos de pacientes: desde los que son completamente autónomos en sus cuidados hasta los que son casi completamente dependientes, incluyendo todas las situaciones intermedias.

Aunque los desarrollos realizados pueden aplicarse a diferentes patologías, el trabajo se ha centrado en pacientes cardiovasculares y cardiometabólicos, para los que son fundamentales tres elementos: el control de la medicación, el control de la actividad física y el control de las constantes vitales. Se trata de un tipo de pacientes con gran incidencia, particularmente en la población de mayor edad y para las que el cuidado es fundamental [4]. Según el estudio presentado en [5] para Estados Unidos, se estima que la prevalencia de enfermedades cardiovasculares es del 48% de los adultos, lo que representa más de 120 millones de personas sólo en ese país. Con respecto al

número de muertes que causan estas enfermedades, los datos presentados en [6] indican que, en 2016, 1 de cada 4 muertes fueron atribuibles a este tipo de enfermedades, lo que representa más de 17 millones de muertes a nivel global.

II. Funcionalidades requeridas

Las principales funcionalidades requeridas por la app, y determinadas tras entrevistas con pacientes, cuidadores y profesionales de la salud, fueron las siguientes:

- Control de la medicación, con alarmas específicas.
- Acceso simple a la información de cada medicamento por los pacientes.
- Control de la actividad física de los pacientes.
- Control de los datos de salud (tensión arterial, peso, etc.) de los pacientes.
- Establecimiento de objetivos tanto para actividad como para datos de salud.
- Posibilidad de realizar llamadas de emergencia (para pacientes dependientes).
- Seguimiento GPS para pacientes dependientes.
- Acceso simple a información de interés para cuidadores.
- Comunicación entre cuidadores (para pacientes que requieren cuidados por turnos).
- Generación de informes periódicos para cuidadores.

A partir de estas funcionalidades requeridas, se planteó el desarrollo de una herramienta de Telemedicina capaz de satisfacer estas demandas. El desarrolló se limitó al sistema operativo Android, dado que el objetivo era crear un prototipo para su posterior validación. Se seleccionó Android por ser el sistema operativo más extendido [7], y por sus ventas comparativas con otros sistemas operativos como iOS [8].

III. Estructura del sistema desarrollado

La estructura del sistema se basa en los siguientes elementos:

1) App móvil, que dispone de una zona para el paciente y una zona para los cuidadores.
2) *Wearables*, que proporcionan información sobre las constantes vitales del paciente a la app móvil.
3) Servidor que gestiona la comunicación con los pacientes y los cuidadores.
4) Base de datos donde se almacena toda la información de pacientes y cuidadores encriptada y anonimizada.

El funcionamiento conjunto de todos los elementos se detalla en la Fig. 1, donde se aprecia cómo a la app pueden acceder tanto los pacientes como los cuidadores; y también como los cuidadores pueden acceder a informes periódicos sobre la situación de los pacientes.

IV. Diseño de la app

El diseño de la aplicación móvil busca la máxima simplicidad de utilización, tanto para los pacientes como para los cuidadores. En particular, los pacientes son habitualmente personas de edad avanzada con escasos conocimientos tecnológicos.

A. Zona de pacientes

La pantalla principal está pensada para el paciente. Muestra información sobre la consecución de sus objetivos de salud, y le permite acceder a tres funcionalidades principales:

- Medicación
- Datos de salud
- Llamada de emergencia

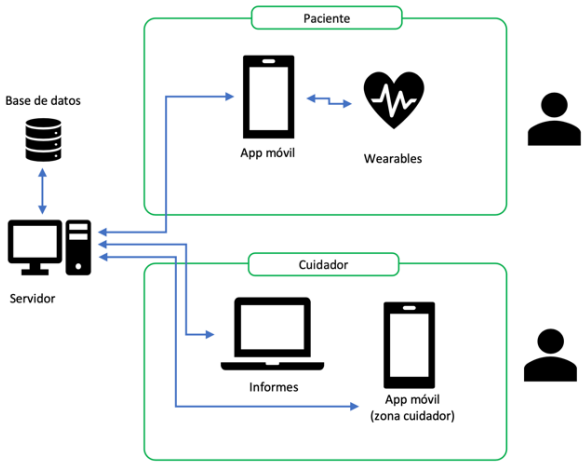


Fig. 1. Estructura del sistema de telemedicina desarrollado en este proyecto.

El aspecto de la pantalla principal se muestra en la Fig. 2. La estructuración de la información en esta pantalla principal está pensada para que el paciente, de un vistazo, pueda comprobar el nivel instantáneo de consecución de sus objetivos (con un indicador porcentual y un icono representativo).

También puede visualizar su progreso, con las estrellas de la parte izquierda que muestran, con diferentes colores, los objetivos logrados el día anterior, así como la semana y el mes anteriores. Adicionalmente, se muestra una frase de ánimo para el paciente en función de su evolución.

El acceso a las tres funcionalidades principales se realiza mediante tres grandes botones en la mitad inferior de la pantalla:

- Comprobación de actividad física y resto de datos de salud, que lleva a una nueva pantalla donde se pueden visualizar, global o independientemente, los valores de datos de salud y actividad física registrados a lo largo del tiempo.
  Estos valores pueden ser comparados con los objetivos establecidos, y pueden mostrarse como un resumen numérico o como un gráfico de evolución.

- Comprobación de medicación y alarmas, que también lleva a una nueva pantalla donde pueden gestionar las pautas de medicación y las alarmas de toma de medicamentos.
  Una funcionalidad adicional permite acceder a información básica sobre cada medicamento mediante el escaneo de su envase.

- Llamada de emergencia, que realiza una llamada a un contacto predefinido simplemente con pulsar el botón.

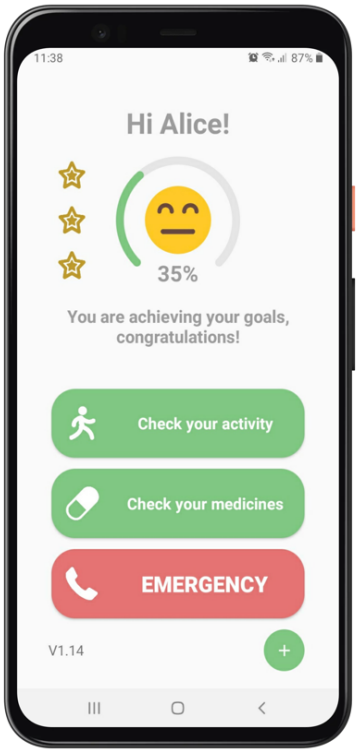


Fig. 2. Pantalla principal, pensada para el paciente.

B. Zona de cuidadores

Desde la pantalla principal, el botón "+", situado en la parte inferior derecha, permite el acceso a la zona de cuidadores. Está pensada para las dos posibilidades más comunes:

- Pacientes *independientes*, que se cuidan a sí mismos.
- Pacientes *dependientes*, que necesitan personal que los cuide.

En el primero de los casos, el botón lleva directamente a la zona de cuidadores (el paciente es su propio cuidador). En el segundo de los casos, es necesario el acceso mediante usuario y contraseña (existe un cuidador externo).

En ambos casos, se llega a una pantalla con un listado de funcionalidades adicionales, que se muestra en las Figs. 3 y 4.

Para facilitar la labor de los cuidadores, sus funcionalidades se agrupan en diversos apartados, que son accedidos mediante una lista expandible.

Las categorías de funcionalidades son:

1. Mis medicinas, donde se gestionan medicamentos y alarmas del mismo modo que desde la pantalla del paciente, pero con ciertas particularidades. En concreto, el establecimiento de pautas de medicación y alarmas está restringido para que sólo puedan realizarlo los cuidadores en caso de pacientes dependientes.

2. Mis datos de salud, que lleva también a la misma funcionalidad a la que acceden los pacientes desde la pantalla principal, pero también con limitaciones. En este caso, el establecimiento de objetivos está restringido sólo para los cuidadores, los pacientes no tienen permisos para modificarlos.

3. Utilidades, que engloba diferentes herramientas que pueden utilizar los cuidadores, según las circunstancias de los pacientes: notas entre cuidadores y seguimiento GPS (serán detalladas más adelante).

4. Información, que comprende dos funcionalidades útiles para cuidadores: vídeos de ayuda sobre cuidados a pacientes y búsqueda de ubicaciones para localizar rápidamente hospitales, farmacias o centros de salud.

5. Gestión, que permite editar la información de los distintos usuarios de la app (paciente y cuidadores), así como notificar incidencias sobre el uso de la app o consultar las condiciones de privacidad.

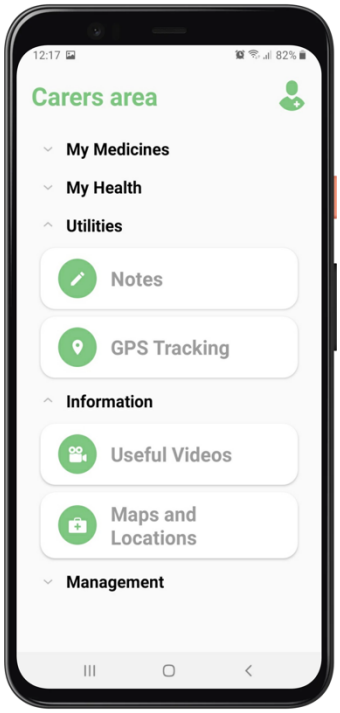


Fig. 3. Funcionalidades de la pantalla de cuidadores con las opciones de Utilidades e Información desplegadas.

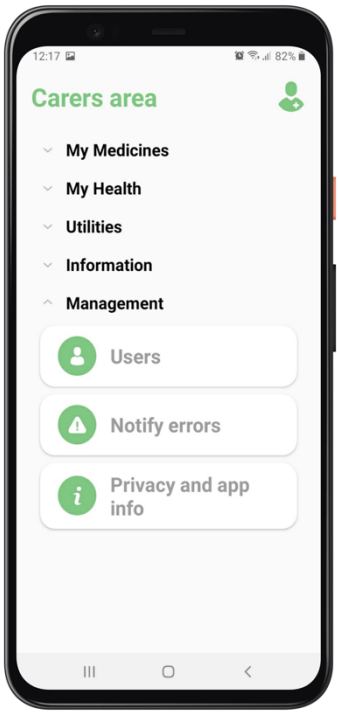


Fig. 4. Funcionalidades de la pantalla de cuidadores con la opción de Gestión activada.

De las funcionalidades mencionadas en la lista anterior, cabe detallar dos de ellas: las notas entre cuidadores y el seguimiento GPS.

Las notas permiten la comunicación entre cuidadores y están diseñadas como un cuaderno en el que la información se ordena temporalmente. Cada cuidador, al comenzar su turno, tiene disponibles las notas introducidas por el cuidador del turno previo, que pueden incluir texto, imágenes o audio, y que permiten ver rápidamente las circunstancias excepcionales que han podido suceder en el turno previo (estado de salud, alimentación, toma de medicamentos, etc.).

El seguimiento GPS está pensado para los pacientes dependientes con enfermedades cognitivas incapacitantes. Los cuidadores pueden activar el seguimiento y obtienen un enlace desde el que pueden acceder a la ubicación del paciente en cualquier momento. Esta funcionalidad está pensada para evitar pérdidas o desorientaciones en ese tipo de pacientes.

V. Herramientas de desarrollo y librerías utilizadas

Tanto la base de datos como el API REST [9] desarrollado para la conexión con la aplicación móvil se han alojado en un servidor Windows Server 2019 [10]. La programación del servidor se ha realizado en PHP, accediendo a bases de datos Microsoft Access.

La aplicación móvil se ha desarrollado para el sistema operativo Android, utilizando el entorno de programación Android Studio [11], versión 4.1.1, para sistema operativo macOS. Además de las librerías estándar para el desarrollo en Android, se han utilizado dos APIs de Google para ciertas funcionalidades:

A. Google Fit API para Android

Se ha utilizado el API de salud de Google [12] para acceder de forma indirecta a los datos de salud de los pacientes. De ese modo, ha sido posible utilizar un mismo procedimiento para obtener información de pacientes que únicamente utilizan su teléfono como para obtener información de pacientes que combinan su teléfono con distintos wearables (pulseras de actividad, básculas, dispositivos de medida de presión arterial, glucómetros, etc.). La interacción con el API de Google Fit se ha realizado de modo bidireccional: la app lee los datos de salud almacenados en Google Fit y también escribe nuevos datos en Google Fit (datos introducidos por los propios pacientes y sus cuidadores). El objetivo es utilizar Google Fit como base de datos externa para el almacenamiento de todos los datos de salud.

B. Google Mobile Vision API

Se ha utilizado el módulo de reconocimiento de texto de esta API, actualmente parte del Kit de Aprendizaje Automático de Google [13] para permitir detectar los medicamentos a partir de las imágenes de sus envases. De este modo, los pacientes pueden obtener información sobre sus medicamentos simplemente escaneando los envases con las cámaras de sus teléfonos. La información que se pone a disposición de los pacientes no es la contenida en los prospectos, sino información personalizada que ha sido introducida previamente por los cuidadores. El funcionamiento es muy simple, pensado para usuarios con pocas habilidades tecnológicas: el paciente apunta con su cámara al envase y, cuando se detecta una coincidencia con uno de los medicamentos prescritos para el paciente, se muestra una pantalla donde es posible acceder a información (en modo audio) relacionada con diversos aspectos: para qué sirve el medicamento, cuándo debe tomarse, cómo debe tomarse, cómo conservarlo, etc.

VI. Conclusiones

Como conclusión, se ha desarrollado una aplicación de telemedicina pensada tanto para ayudar a pacientes como para ayudar a sus cuidadores. La aplicación cubre las principales funcionalidades demandadas por ambos colectivos, de acuerdo con la información recopilada en contactos previos con ellos. Las pruebas internas realizadas hasta la fecha muestran un correcto funcionamiento para la app en sus diversas funcionalidades. Los trabajos futuros incluyen la realización de pruebas con usuarios reales y la búsqueda de socios para una posible comercialización.

Agradecimientos

Este trabajo ha sido posible realizarlo gracias al proyecto PROMETEU/2017/173 financiado por la Consellería de Educación, Investigación y Deporte de la Comunidad Valenciana.

Referencias

[1] J.J., Mira, "Tecnologías móviles e inalámbricas para cuidar nuestra salud", *Journal of healthcare quality research*, *33*(4), 183-186, 2018

[2] K. Cheikh-Moussa, J.J. Mira & D. Orozco-Beltran, "Improving Engagement Among Patients With Chronic Cardiometabolic Conditions Using mHealth: Critical Review of Reviews", *JMIR Mhealth Uhealth,* 8(4):e15446, 2020.

[3] V. Pérez-Jover, M. Sala-González, M. Guilabert & J.J. Mira, "Mobile apps for increasing treatment adherence: systematic review"., *JMIR*, *21*(6), e12505, 2019.

[4] N. Sattar, J.M.R. Gill & W. Alazawi, "Improving prevention strategies for cardiometabolic disease." *Nat Med* **26,** 320–325, 2020. https://doi.org/10.1038/s41591-020-0786-7

[5] S.S. Khan, S. Sidney, D.M. Lloyd-Jones & J.S. Rana, "National and Global Trends of Cardiovascular Disease Mortality, Morbidity, and Risk", *ASPC Manual of Preventive Cardiology*, pp. 17-33, Springer, Cham, 2021

[6] Read, S. H., & Wild, S. H. (2020). Prevention of premature cardiovascular death worldwide. *The Lancet*, *395*(10226), 758-760.

[7] *Mobile operating systems' market share worldwide from January 2012 to January 2021*. https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/

[8] R. Győrödi,D. Zmaranda, V.G. Adrian & C. Győrödi, "A comparative study between applications developed for Android and iOS", *International Journal of Advanced Computer Science and Applications*, *8*(11), 176-182, 2017

[9] M. Masse, "REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces", *O'Reilly Media, Inc.,* 2011

[10] *Microsoft Windows Server*. https://www.microsoft.com/es-es/windows-server

[11] *Android Studio*. https://developer.android.com/studio

[12] *Google Fit APIs*. https://developers.google.com/fit

[13] *Machine Learning Kit and Google Vision APIs*. https://developers.google.com/ml-kit


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Detección temprana de cyberbullying en redes sociales

Manuel F. López-Vizcaíno, Francisco J. Nóvoa, Víctor Carneiro, Fidel Cacheda
Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC)
Departamento de Ciencias de la Computación y Tecnologías de la Información
Universidade da Coruña
Campus de Elviña, A Coruña, 15071
manuel.fernandezl@udc.es, francisco.javier.novoa@udc.es, victor.carneiro@udc.es, fidel.cacheda@udc.es

En este resumen presentamos dos enfoques que tienen en cuenta el tiempo en la detección del cyberbullying en redes sociales. Siguiendo un método de aprendizaje máquina supervisado, definimos dos modelos específicos de detección temprana, denominados umbral y dual. El primero sigue un enfoque más simple, mientras que el segundo requiere dos modelos de aprendizaje automático. Como contribuciones principales destacar la definición de dos grupos de características y dos métodos de detección temprana, diseñados específicamente para este problema. Realizamos una evaluación pormenorizada utilizando un conjunto de datos real y siguiendo una evaluación basada en el tiempo que penaliza las detecciones tardías. Nuestros resultados demuestran que somos capaces de mejorar los modelos de detección de referencia hasta un 42%.

***Palabras Clave*—cyberbullying, ciberacoso, redes sociales, detección temprana, aprendizaje máquina**

I. RESUMEN

Este artículo presenta un resumen de la investigación detallada en [1]. En este trabajo, las principales contribuciones fueron:

- Definir y caracterizar el problema de la detección temprana de cyberbullying en redes sociales.
- Presentar dos modelos de aprendizaje máquina (umbral y dual) y dos tipos de características (conjuntos de palabras y temporales), específicos para este problema.
- Realizar experimentos exhaustivos usando un dataset real y siguiendo una evaluación basada en el tiempo que demuestran las mejoras en el rendimiento de los modelos propuestos.

El problema de la detección de temprana de cyberbullying consiste en, dada una sesión en una red social, determinar si dicha sesión se corresponde con cyberbullying procesando el menor número posible de posts. Para el estudio de este problema se ha utilizado un dataset público basado en la red social Vine.

A partir del dataset se han extraído las siguientes características básicas: perfil del propietario de la sesión, indicadores de la sesión, características de los comentarios, características del vídeo, características LDA (Latent Dirichlet Allocation). Además, este conjunto de características se expande con: similaridad entre comentarios usando una representación de conjuntos de palabras y aspectos temporales.

En lo que respecta a los modelos propuestos, el modelo umbral integra una función de decisión basada en la probabilidad de clase que determina si existen evidencias suficientes para tomar una decisión final. Por otra parte, el objetivo del modelo dual es determinar de manera independiente cada opción (es decir, cyberbullying o no), utilizando dos modelos de aprendizaje máquina y dos funciones de decisión, independientes en ambos casos.

Los resultados de nuestros experimentos demuestran que el modelo umbral mejora significativamente los resultados de los modelos de referencia hasta un 26% y el modelo dual consigue mejoras de hasta un 42%. Además, la incorporación de las características propuestas a las características básicas permite obtener el mejor rendimiento en ambos modelos.

AGRADECIMIENTOS

Esta investigación ha sido financiada por el Ministerio de Economía y Competitivad de España y fondos FEDER de la Unión Europea (Proyecto PID2019-111388GB-I00) y por el Centro de Investigación de Galicia "CITIC", financiado por la Xunta de Galicia y la Unión Europea (programa FEDER GALICIA 2014-2020), mediante la ayuda ED431G 2019/01.

REFERENCIAS

[1] López-Vizcaíno, Manuel F., Francisco J. Nóvoa, Victor Carneiro, and Fidel Cacheda. "Early detection of cyberbullying on social media networks." Future Generation Computer Systems 118 (2021): 219-229.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Dataset anotado para detección de anomalías en un CPD con sensores IoT

Laura Vigoya, Diego Fernández, Victor Carneiro, Fidel Cacheda
Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC)
Departamento de Ciencias de la Computación y Tecnologías de la Información
Universidade da Coruña, A Coruña, España
l.v.vigoya@udc.es, dfernandezi@udc.es, victor.carneiro@udc.es, fidel.cacheda@udc.es

Para detectar anomalías de tráfico usando técnicas de aprendizaje automático, es necesario obtener un conocimiento básico del rendimiento y comportamiento de la red. Por ello este trabajo presenta DAD, un conjunto de datos contextualizado, etiquetado y con un análisis sistemático, que reproduce el comportamiento de una red real. DAD presenta siete días de actividad con tres tipos de anomalías: duplicación, interceptación y modificación. Sobre el conjunto de datos se aplica ingeniería de características, para que así puedan ser aplicadas técnicas de clasificación usando algoritmos de aprendizaje automático.

***Palabras Clave*—dataset; IoT; MQTT; sensores; ingeniería de características**

I. INTRODUCCIÓN

Internet de las Cosas (IoT) ha cobrado especial relevancia debido a la rápida aparición de nuevos entornos de interconexión, donde cada vez más dispositivos se conectan a una red con requerimientos que hacen que la eficiencia, el bajo coste, la eficiencia, la tolerancia a fallos o la flexibilidad adquieran cada vez mayor importancia. Sin embargo, estos nuevos escenarios son también susceptibles de sufrir ataques. Por tanto, paliar las posibles consecuencias de dichos ataques se convierte en un reto. Para poder mitigar estos ataques y detectar posibles anomalías, es necesario comprender el comportamiento de las redes IP.

En este trabajo se presenta un conjunto de datos etiquetado, DAD, obtenido a partir de los datos proporcionados por los sensores de temperatura inteligente del Centro de Procesamiento de Datos (CPD) del Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC) durante una semana. Sobre estos datos, se aplican técnicas de aprendizaje automático con el fin de detectar posibles anomalías en el tráfico [1].

En el escenario aquí propuesto se hace uso del protocolo Message Queuing Telemetry Transport (MQTT).

II. DATASET

La infraestructura virtual necesaria para la creación del dataset está formada por cinco máquinas virtuales, un broker MQTT y cuatro clientes. Para aproximar el dataset a un entorno real, los datos fueron obtenidos de los sensores de las InRow del CPD indicado. Estos sensores están conectados a una red IoT interna.

El modelado matemático de los sensores se llevó a cabo haciendo uso de series de tiempo. El método seleccionado para generar los datos de temperatura fue un método de descomposición STL.

DAD tiene un total de 101.583 paquetes, 3.4% UDP y 96.9% TCP. El 63.3% del total son paquetes MQTT y el 16% de estos paquetes son anomalías. Los datos de temperatura se encuentran en la carga útil del mensaje MQTT. Se tuvieron en cuenta tres tipos de anomalías: duplicación, interceptación y modificación.

Además del análisis de paquetes, se realizó un análisis de flujos. Los flujos se establecen como unidireccionales, usando como umbral de inactividad 30 segundos. El dataset tiene un total de 67.848 flujos, de los cuales 544 corresponden a flujos anómalos.

III. CONCLUSIONES

En este artículo se presenta DAD, un dataset con tráfico IoT etiquetado. También se realiza un análisis de características que simplifica la posible aplicación de algoritmos para la detección de tráfico anómalo.

AGRADECIMIENTOS

Este trabajo fue parcialmente financiado por el Ministerio de Ciencia e Innovación, Plan Nacional de Investigación y Desarrollo (proyecto PID2019-111388GB-I00).

REFERENCIAS

[1] Vigoya, Laura, Fernández, Diego, Carneiro, Victor y Cacheda, Fidel, "Annotated Dataset for Anomaly Detection in a Data Center with IoT Sensors", Sensors, vol. 20, 2020, DOI=10.3390/s20133745

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Reconocimiento de emociones para la mejora de la seguridad en la conducción

Ignacio Prieto, Víctor Corcoba, David Melendi, Laura Pozueco, Xabiel G. Pañeda, Roberto García
Departamento de Informática,
Universidad de Oviedo
Campus de Xixón, sn, 33203, Xixón, Asturias, España
{UO252814, corcobavictor, melendi, pozuecolaura, xabiel, garciaroberto}@uniovi.es

En este trabajo se presenta un sistema de reconocimiento de emociones orientado a mejorar la seguridad en la conducción. El sistema captura y analiza imágenes del conductor para determinar su estado anímico. Para ello, se ha desarrollado un prototipo que muestra al usuario las emociones que detecta, infiriendo su estado mediante la utilización de mecanismos de aprendizaje automático. Durante el desarrollo del prototipo se han utilizado distintos modelos de aprendizaje automático. Se ha observado que los modelos generalistas pre-entrenados no proporcionan buenos resultados, siendo mucho más adecuados los modelos ad hoc. Durante el desarrollo del proyecto se han realizado pruebas de concepto en un entorno controlado. El objetivo es el de integrar el prototipo en un sistema avanzado de asistencia al conductor (ADAS). Con una visión holística de la seguridad en la conducción, este ADAS integrará la información del conductor, del vehículo y del entorno para ofrecer recomendaciones al usuario.

***Palabras Clave*- comportamiento de los conductores, visión por computador, sistema avanzado de asistencia al conductor**

I. INTRODUCCIÓN

En los últimos años se ha experimentado una evolución tecnológica notable en ámbitos que, hasta la fecha, habían tenido un impacto moderado en la sociedad. Hoy en día, ya es frecuente encontrar productos que integran algún tipo de inteligencia artificial (IA) o dispositivos que podemos enmarcar dentro del denominado Internet de las Cosas (IoT). Uno de los sectores en los que se han implantado este tipo de productos innovadores es el de la automoción.

En el sector de la automoción se han introducido numerosas mejoras tecnológicas, entre las que podemos destacar una importante implantación de sistemas avanzados de asistencia al conductor o *Advanced Driver Assistance Systems* (ADAS). Algunos ejemplos son los controles adaptativos de velocidad de crucero, los sistemas de alerta de colisión con peatones y ciclistas o los sistemas de advertencia de colisión frontal. Dado que un alto porcentaje de los accidentes de tráfico se produce principalmente debido al comportamiento de los conductores [1][2], tienen especial importancia aquellos ADAS que de algún modo u otro intentan determinar el estado de los conductores con la finalidad de evitar accidentes. Entre estos ADAS se encuentran, por ejemplo, los sistemas de detección de fatiga o *Driver Fatigue Detectors* (DFD), que se han ido implantando progresivamente en los vehículos modernos. Estos sistemas monitorizan el estado del conductor y generan algún tipo de advertencia que recomienda detener el vehículo en caso de considerar que el estado del piloto no es apto para la conducción.

Algunos estudios recientes resaltan la relación entre nuestras emociones y los accidentes, e indican que sólo siete de cada cien conductores mantienen un control emocional correcto y equilibrado al volante [3]. Estos estudios indican que las emociones influyen directamente en el riesgo de sufrir un accidente. Por ello, un buen mecanismo para mejorar la seguridad en la conducción puede ser el de realizar recomendaciones mediante el análisis del estado anímico de los conductores.

En este trabajo se presenta un prototipo de un sistema *edge computing* que pretende establecer el estado anímico del conductor a partir de la monitorización de su rostro. El sistema está dotado de una cámara enfocada hacia el conductor y, utilizando técnicas de IA, determina si éste está enfadado, sorprendido, feliz, etc. Para ello, el sistema se basa en la utilización de una red neuronal que ha sido entrenada previamente con un banco de imágenes que contiene 35.888 ficheros de 48x48 píxeles [4] El prototipo presentado en este trabajo solamente muestra la emoción detectada. No obstante, el objetivo final es el de integrar este sistema dentro de un ADAS complejo que, con una visión holística, integre el estado del conductor, del vehículo y del entorno en un sistema de recomendaciones con vistas a mejorar la eficiencia y la seguridad en la conducción [5].

El resto del trabajo se ha estructurado como sigue. En la Sección II se realiza un breve resumen de trabajos anteriores. En la Sección III se proporcionan detalles acerca de la implementación del sistema desarrollado. En la Sección IV se presentan los resultados obtenidos con cada uno de los modelos de inferencia utilizados durante las pruebas. En la Sección V se describen algunas pruebas de concepto realizadas con el prototipo. Finalmente, en la Sección VI se incluyen las conclusiones y una descripción de los trabajos futuros.

II. Trabajos Relacionados

Los ADAS que pretenden mejorar la seguridad en la conducción mediante la observación de los conductores, como los DFD, utilizan básicamente tres técnicas [6]:

- Análisis del estado del conductor, mediante la detección de cambios fisiológicos.
- Análisis del comportamiento del conductor, mediante la monitorización de aspectos del vehículo como su posición y trayectoria.
- Análisis combinado del estado del conductor y de su comportamiento.

El análisis del estado del conductor se realiza principalmente observando los movimientos de la cabeza, de los ojos y de los párpados mediante el uso de técnicas de visión por computador. Un ejemplo es el sistema AntiSleep de Smart Eye AB [7]. Se ha discutido mucho sobre la conveniencia de utilizar este tipo de monitorización, debido a que muchos de estos sistemas miden los efectos de la fatiga en los conductores y no la fatiga en si misma [8], por lo que es imposible anticiparse para evitar un eventual accidente. Por ello, en otros trabajos anteriores se baraja la posibilidad de realizar mediciones sobre otros órganos como la piel [9] o el corazón [10] o de utilizar de forma combinada imágenes del conductor e información del vehículo [11]. No obstante, la monitorización de la cara de los conductores no tiene por qué limitarse a estimar su fatiga, sino que puede extenderse a otro tipo de situaciones. Esta es, por ejemplo, la propuesta de [12], en el que los autores se centran en la detección de distracciones durante la conducción.

La detección de expresiones faciales es otro campo en el que se ha trabajado intensamente los últimos años. En [13] los autores incluyen una extensa revisión de trabajos anteriores que exploran la identificación de expresiones faciales. Se comparan distintas bases de datos de referencia para el entrenamiento de sistemas de aprendizaje automático, entre las que se encuentra la utilizada en este trabajo [4] Adicionalmente, se presentan los pasos básicos para la identificación de expresiones faciales y una revisión de redes neuronales diseñadas para este fin. Algunas de estas estructuras se comparan en este trabajo. [14][15].

III. Implementación del Prototipo

A. Elementos hardware e interfaz de usuario

El sistema se ejecuta en un ordenador en placa (SBC) al que se conecta una webcam convencional para capturar las imágenes del conductor. En el prototipo que se ha desarrollado, se ha utilizado una Raspberry Pi 4 y una webcam Joyaccess X-WB01FR.

Para mostrar al conductor las emociones registradas por el sistema, se utiliza un panel led flexible de 8x32 píxeles. El panel se ha conectado a los puertos de propósito general de entrada/salida del SBC y se controla a través de la API *neopixel*. Se han desarrollado unas librerías que, actuando sobre esta API, escriben en el panel led las emociones inferidas por el sistema utilizando cadenas de texto.

Todos los componentes se han integrado en una carcasa elaborada al efecto, que en su parte delantera dispone de una lámina plástica protectora con dos capas adhesivas de tinte oscuro. La intención final del prototipo es la de disponer de un dispositivo embarcado. El resultado final se puede observar en la Figura 1.

B. Detección del rostro del conductor y estimación de emociones en el dispositivo embarcado

Respecto al dispositivo embarcado, el sistema que se ejecuta en la placa SBC utiliza la librería OpenCV para detectar las caras de los conductores en las imágenes que se capturan con la webcam. Se usan clasificadores en cascada basados en funciones de Haar [16].

Para adecuar las imágenes capturadas al modelo utilizado para la inferencia, las caras detectadas se transforman a escala de grises y se reduce su resolución a 48x48 píxeles (se pasan a un formato similar al utilizado durante la fase de entrenamiento del modelo).

Posteriormente, las caras se pasan a un modelo implementado en el SBC utilizando las librerías de aprendizaje automático de TensorFlow. Más concretamente, se ha utilizado la API de alto nivel de TensorFlow, Keras.

Fig. 1. Prototipo final integrado

Si la inferencia se ha realizado con un nivel de certidumbre superior a un umbral configurable, la emoción se pinta en el panel led del prototipo. Esta retroalimentación se eliminará en una futura integración del sistema en un ADAS, siendo sustituida por algún tipo de recomendación relativa a la seguridad de la conducción.

Para que la inferencia se pueda realizar en un dispositivo embarcado de capacidad de cómputo limitada,

el entrenamiento del modelo implementado con TensorFlow se realiza fuera de línea utilizando el banco de imágenes que se describe en el punto siguiente. Durante este proceso de entrenamiento, se genera un fichero *.hdf5* que contiene toda la parametrización necesaria. Este fichero se carga posteriormente en el SBC para realizar la inferencia en línea.

C. Banco de imágenes de entrenamiento

El sistema funciona infiriendo el estado anímico del conductor a partir de una imagen de su cara. Para entrenar al modelo que tiene que realizar esta inferencia, es conveniente utilizar un banco de imágenes preclasificadas con un volumen que permita un entrenamiento efectivo. Existen varios conjuntos de datos de este tipo, entre los que caben destacar el TFD [17] con 112.234 imágenes, Multi-PIE [18] con 755.370 imágenes, EmotioNet [19] con 1 millón de imágenes, RAF-DB [20] con 29.672 imágenes, AffectNet [21] con 450.000 imágenes o ExpW [22] con 91.793 imágenes. En nuestro caso hemos utilizado el banco de imágenes FER2013 [4], que ya se ha utilizado como base para entrenar sistemas de reconocimiento de expresiones faciales [13]. Este banco de imágenes contiene 35.888 fotos en blanco y negro de caras de personas con una resolución de 48x48 píxeles. Se trata de imágenes recogidas automáticamente por la API del buscador de imágenes de Google. Han sido registradas después de un preprocesamiento que incluye la normalización de las características de las imágenes y la eliminación de imágenes etiquetadas incorrectamente. Cada una de estas imágenes está asociada a una de siete emociones:

- 0: Enfado
- 1: Repulsión
- 2: Asustado
- 3: Felicidad
- 4: Tristeza
- 5: Sorpresa
- 6: Neutro

La disponibilidad de fotos en el banco de imágenes para cada una de estas emociones es desigual, tal y como se observa en la Figura 2. Un ejemplo de imágenes para una de estas emociones se muestra en la Figura 3.

Siguiendo las recomendaciones de los autores del banco de imágenes, un 80% de ellas se utilizarán para el entrenamiento del modelo, reservando el 20% restante para pruebas de validación.

D. Inferencia de emociones y entrenamiento del modelo

Como se comentó en el punto anterior, el modelo se entrena fuera de línea. Para ello, se han realizado unos programas usando las librerías de TensorFlow. Más concretamente, se ha utilizado la API de alto nivel de TensorFlow, Keras. En concreto, se usa un modelo secuencial basado en varias capas.

Estos programas se han ejecutado sobre ordenadores equipados con tarjetas gráficas avanzadas. En concreto, se han realizado pruebas con un ordenador con una tarjeta NVIDIA GEFORCE RTX 2070 SUPER con 2560 núcleos CUDA y con otro ordenador equipado con una tarjeta NVIDIA GEFORCE RTX 3070 con 5888 núcleos CUDA. En ambos casos los componentes utilizados han sido Python, Tensorflow, Keras, CUDA y la biblioteca cuDNN. Debido a restricciones de compatibilidad, fue necesario trabajar con versiones diferentes para cada uno de estos entornos. En concreto, con la tarjeta de la generación 20 se utilizó Python 3.7, Tensorflow 2.1, CUDA 10.0 y cuDDN 7.0. Por otro lado, la generación 30 presenta una arquitectura Ampere, lo que obliga estrictamente al uso de CUDA 11.1 y cuDNN 8.0. Esto a su vez implica utilizar Python 3.8 y Tensorflow 2.5. Aunque se realizaron pruebas de entrenamiento con las CPU de los ordenadores, los tiempos requeridos pasan de unas horas a varios días.

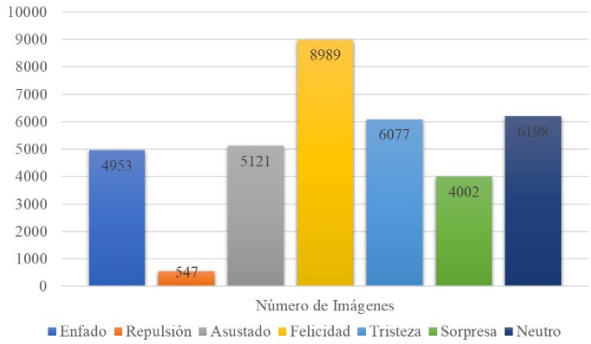


Fig. 2. Número de imágenes del banco de imágenes por emoción

Fig. 3. Ejemplo de imágenes etiquetadas como “sorpresa”

En el desarrollo del sistema de inferencia se han utilizado varios modelos, pudiendo comparar sus resultados. En las primeras versiones se emplearon modelos VGG [14] pre-entrenados. Se trata de redes neuronales convolucionales muy profundas que presentan un conocimiento genérico para detectar patrones en imágenes. Las últimas capas de estos modelos son muy simples, por lo que se han personalizado durante el desarrollo. En las últimas versiones, se usaron modelos CNN [15], redes neuronales convolucionales que, aunque emplean un funcionamiento similar a los VGG, admiten una mayor personalización. En concreto, se han utilizado los siguientes modelos:

- Modelo VGG16 y red neuronal de 3 capas.
- Modelo VGG16 y red neuronal de 1 capa.
- Modelo VGG19 y red neuronal de 3 capas.
- 4 modelos CCN. El último modelo CNN incluye una reducción de categorías. Se realizaron varias reducciones de categorías, obteniendo los mejores resultados con las cuatro emociones más

relevantes para el estudio (0: "Enfado", 3: "Felicidad", 5: "Sorpresa" y 6: "Neutro")

En la Tabla I se muestran los resultados obtenidos con cada uno de estos modelos, que se describen de forma más detallada en la sección III.

Finalmente, para poder monitorizar el proceso de entrenamiento se ha utilizado el sistema Tensorboard. Este sistema ha permitido observar, en tiempo real, la evolución de los modelos a medida que se avanzaba en las épocas.

Tabla I
RESULTADOS OBTENIDOS CON CADA MODELO Y CONJUNTO DE DATOS

	Datos de entrenamiento		Datos de validación	
Modelo	Precisión máxima	Época	Precisión máxima	Época
VGG16 y red de 3 capas	0,8009	2434	0,4749	263
VGG16 y red de 1 capa	0,8359	1807	0,4948	650
VGG19 y red de 3 capas	0,71	1362	0,4593	323
CCN #1	0,99	148	0,6050	794
CCN #2	0,6829	148	0,6503	119
CCN #3	0,7389	297	0,702	297
CCN #4	0,869	277	0,853	277

IV. MODELOS DE INFERENCIA UTILIZADOS

A. Modelo pre-entrenado VGG16 y red neuronal de 3 capas.

El modelo VGG16 posee 16 capas, donde las 3 últimas representan una red neuronal simple. En el prototipo implementado, se sustituyeron estas tres últimas capas por una red neuronal diseñada de forma particular para poder ser entrenada de forma específica con el banco de imágenes descrito con anterioridad.

Durante el entrenamiento, se observa una tendencia logarítmica en la precisión del modelo. Tras 263 épocas, la precisión comienza a estabilizarse. En cualquier caso, tal y como se observa en la Tabla I los resultados con este modelo no son buenos. La precisión sobre el conjunto de validación se estabiliza en un valor bajo, inferior a 0,5 que decrece ligeramente con el trascurso de las épocas. Este efecto refleja un evidente *overfitting*. Por otro lado, en la Figura 4 se puede apreciar que los aciertos para cada emoción, representados en la diagonal principal, son en su mayoría inferiores al 50 %. Las dudas que se plantean para cada clase carecen de relación lógica. Por ejemplo, para un rostro enfadado se generarán dudas considerables con casi todas las demás clases. Esto muestra que el modelo apenas tiene conocimientos para generalizar.

B. Modelo pre-entrenado VGG16 y red neuronal de 1 capa.

Debido a los resultados obtenidos con el modelo VGG16 con una red neuronal particularizada de 3 capas, se decidió simplificar estas capas personalizadas para comprobar si se podía incrementar la precisión y reducir el efecto de *overfitting*.

Los resultados obtenidos son similares a los obtenidos con el modelo anterior, produciéndose una ligera mejora en la precisión con el conjunto de datos de validación, tal y como se observa en la Tabla I.

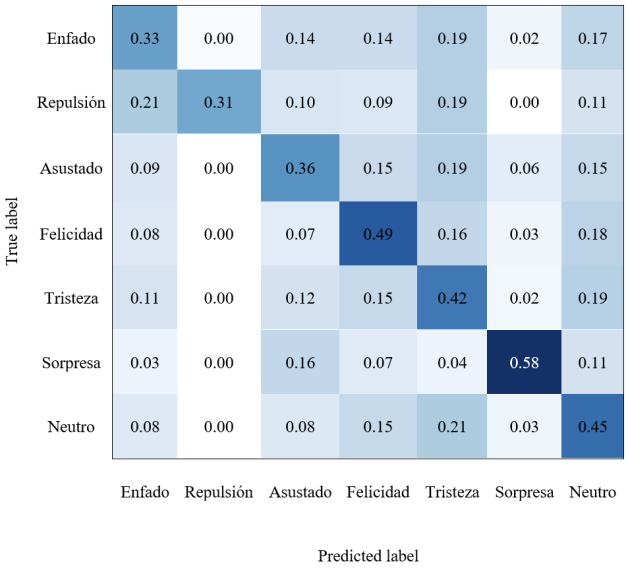


Fig. 4. Matriz de confusión del modelo VGG16 y una red de 3 capas

Nuevamente, se produce un claro efecto de *overfitting*, llegando a conseguir una precisión de 0,8359 en los datos de entrenamiento en tan solo 1807 épocas, pero fallando en la identificación de imágenes en el conjunto de validación.

La conclusión es que el modelo VGG16 no se ajusta correctamente a las necesidades del sistema desarrollado. El modelo pre-entrenado aporta filtros demasiado genéricos que son eficaces para detectar formas de animales o personas. No obstante, no es capaz de captar de forma correcta las expresiones de los rostros.

C. Modelo pre-entrenado VGG19 y red neuronal de 3 capas.

Para observar si se aportaba alguna mejora empleando otro modelo pre-entrenado, se hicieron las pruebas usando VGG19 con una estructura similar a la comentada en el punto III.A. Este modelo incluye 3 capas adicionales de convolución.

Los resultados son similares a los obtenidos con los modelos anteriores. Las cifras de precisión son incluso peores, tal y como se puede observar en la Tabla I. Adicionalmente, se añadió un incremento de la probabilidad para la función de *dropout*. Esto produce una mayor desactivación de neuronas consiguiendo mitigar el efecto de *overfitting*. Así, en las primeras épocas se observan pendientes menos pronunciadas para los datos de entrenamiento. Esto es indicativo de que el uso de la función de *dropout* produce resultados. No obstante, en pocas épocas el modelo vuelve a mostrar una tendencia al *overfitting*.

D. Modelo CCN número 1

Para en este modelo se ha diseñado de forma específica la estructura secuencial de la red. Tal y como se puede observar en la Tabla II, el diseño del modelo incluye tres capas convolucionales, una capa densa de neuronas y una capa de salida. El orden habitual de las capas es aplicar una capa convolucional y después añadir la operación *MaxPooling*.

Tabla II
ESTRUCTURA DEL MODELO CNN NÚMERO 1

	Operaciones	Número	Tamaño
Capa 1	Convolución MaxPooling	32	Filtro(3,3) (2,2)
Capa 2	Convolución MaxPooling	64	Filtro(3,3) (2,2)
Capa 3	Convolución MaxPooling	128	Filtro(3,3) (2,2)
	Flatten		
Capa 1	1024 neuronas Dropout	Activación ReLU 0,5	
Capa salida	7 neuronas	Activación Softmax	

Tal y como se observa en la Tabla I, los resultados obtenidos son mejores que con los modelos VGG. Se ha logrado una gran mejoría en la precisión con los datos de validación, que alcanza un 0,6050. No obstante, se observa un efecto de *overfitting* a partir de la época 40.

A partir de la época 40 se observa un incremento notable en la función de error con los datos de validación acompañado de un lento crecimiento en la precisión. Lo ideal para para evitar el efecto de *overfitting* es detener el entrenamiento en esa época. Por ello, los valores de precisión aproximados que se pueden obtener son del 60% para los datos de validación y de un 70 % para los datos de entrenamiento.

Tal y como se observa en la Figura 5, el modelo presenta valores razonables para algunas de las categorías. Igualmente, las confusiones producidas siguen una tendencia lógica, dada la ambigüedad de algunas imágenes del conjunto de datos.

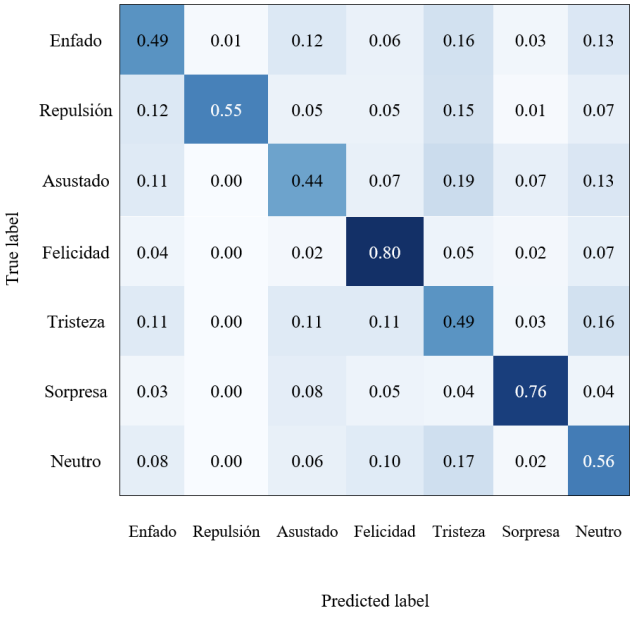


Fig. 5. Matriz de confusión del primer modelo CCN

E. *Modelo CCN número 2*

En este modelo se mantiene la estructura del modelo anterior. No obstante, se ha añadido un paso previo de pre-procesamiento con la finalidad de enriquecer el conjunto de datos disponible. Para ello, se han creado nuevas imágenes introduciendo pequeñas modificaciones en las disponibles en el conjunto de datos de partida, permitiendo asociar a cada estado anímico un mayor número de imágenes.

La mayor parte del conjunto de datos de partida, muestra imágenes centradas y con ángulos similares. Por ello, aplicar un pre-procesamiento generando nuevas imágenes mediante la aplicación de giros, desplazamientos o cambios de perspectiva puede suponer una mayor eficiencia no solo al analizar los datos de validación, sino también en el uso real del sistema.

Tal y como se observa en la Tabla I, al aplicar este pre-procesamiento al conjunto de datos de entrenamiento se consigue mejorar la precisión un 5% aproximadamente respecto al modelo anterior. Adicionalmente, las funciones de error muestran una tendencia decreciente con las épocas, sin observar problemas de *overfitting*.

En la Figura 6 se observa que la precisión ha aumentado en casi todas las categorías. No obstante, la categoría "1: Repulsión" presenta algunas dificultades. Esto puede deberse al desequilibrio que hay en el conjunto de datos de prueba que se muestra en la Figura 1.

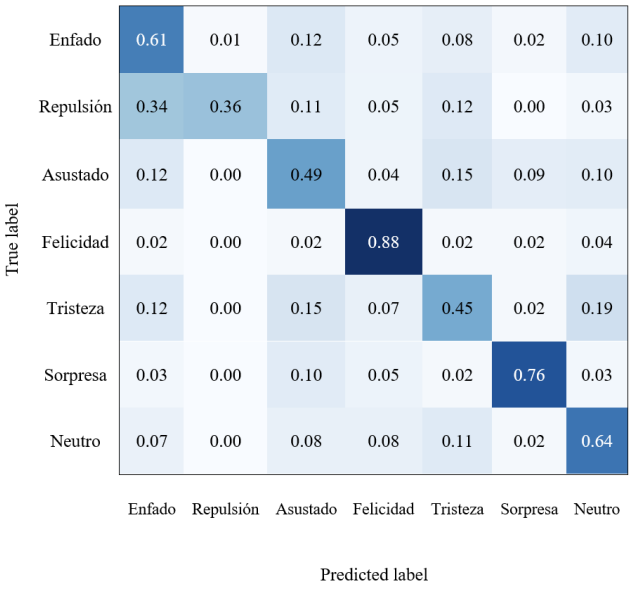


Fig. 6. Matriz de confusión del segundo modelo CCN

F. *Modelo CCN número 3*

Para en este nuevo modelo se ha diseñado una nueva estructura de red neuronal, que se muestra en la Tabla III. Además de aplicar los pre-procesamientos que se utilizaron en el modelo anterior, se ha añadido una capa convolucional adicional, se han incluido funciones *dropout* en las capas convolucionales y se ha añadido otra capa densa.

Tal y como se observa en la Tabla I, los resultados para este modelo son muy buenos, con una mejora casi del 5%

respecto a la versión anterior. La precisión más alta alcanzada se logra en la época 297, logrando un 0,702 para los datos de validación y un 0,7389 para los de entrenamiento.

Por otro lado, el *overfitting* se mantuvo controlado en todo momento. La función de error decrece durante muchas más épocas que en el resto de modelos empleados. Al alcanzar aproximadamente las 150 épocas la función de error se mantiene constante prácticamente.

Nuevamente, en la Figura 7 se observa que la precisión ha aumentado en casi todas las categorías. Emociones como las de felicidad o sorpresa muestran una precisión excelente, del 89% y 86% respectivamente.

Tabla III
ESTRUCTURA DEL MODELO CNN NÚMERO 3

	Operaciones	Número	Tamaño
Capa 1	Convolución	32	Filtro(3,3)
	Dropout		0,25
Capa 2	Convolución	64	Filtro(3,3)
	MaxPooling		(2,2)
	Dropout		0,25
Capa 3	Convolución	128	Filtro(3,3)
	MaxPooling		(2,2)
	Dropout		0,25
Capa 4	Convolución	128	Filtro(3,3)
	MaxPooling		(2,2)
	Dropout		0,25
	Flatten		
Capa 1	2048 neuronas		Activación ReLU
	Dropout		0,5
Capa 2	1024 neuronas		Activación ReLU
	Dropout		0,5
Capa salida	7 neuronas		Activación Softmax

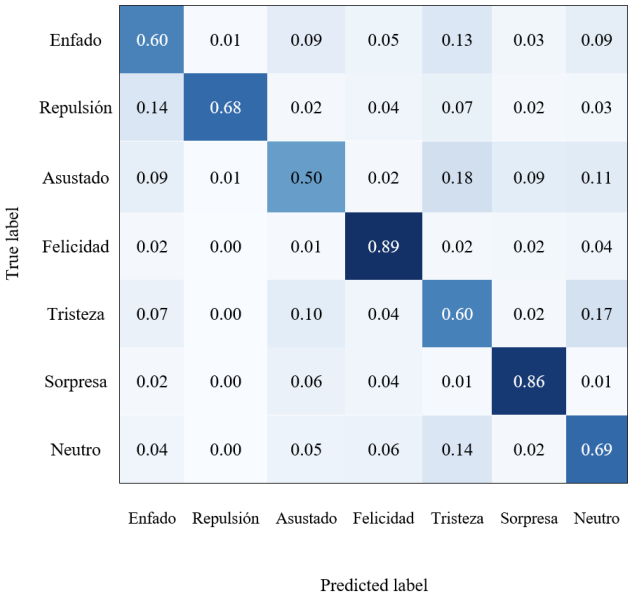


Fig. 7. Matriz de confusión del tercer modelo CCN

G. *Modelo CCN número 4 con reducción de categorías.*

Para comprobar la influencia del conjunto de datos de partida sobre la eficiencia del modelo, se efectuaron pruebas alterando las categorías utilizadas. Para ello, se utilizó la misma estructura que se usó en el punto anterior y que se detalla en la Tabla III.

Como ya se observó en las pruebas anteriores, existen ciertas emociones que son complejas de determinar en imágenes que pueden resultar ambiguas. Mediante la función *Softmax*, para cada imagen, la red proporciona la probabilidad asociada a cada categoría. Es decir, muestra qué porcentaje de cada emoción encuentra en una imagen. No obstante, como el conjunto de datos de partida describe únicamente una etiqueta para cada imagen, estos intentos de mostrar varias emociones son rechazados y castigados por la función de error.

Por todo lo anterior, se realizaron diferentes pruebas prescindiendo de algunas de las categorías existentes en el conjunto de imágenes de partida. El objetivo es ver cómo afectan estas variantes a las predicciones. Si el modelo encuentra, por ejemplo, la emoción de tristeza muy relacionada con la de miedo, eliminando una de ellas se observarán mejoras que, por otro lado, resultan lógicas. Aunque se realizaron diferentes pruebas, solo describimos aquí la prueba final que se hizo reduciendo el modelo a 4 emociones: 0: "Enfado", 3: "Felicidad", 5: "Sorpresa" y 6: "Neutro".

Tal y como se observa en la Tabla I, los resultados obtenidos con 4 emociones son muy positivos. Al eliminarse las categorías más difusas y mantenerse las más claras, se obtiene una gran mejora en la clasificación. Durante la clasificación de los datos de validación se observa que se alcanza una precisión superior al 85% en la época 277, manteniéndose constante en épocas posteriores. Si se extiende el entrenamiento más allá de esa época, se observa un leve efecto de *overfitting*, que comienza a ser evidente a partir de la época 300.

En la Figura 8 se muestra la matriz de confusión obtenida en este modelo. En ella se observa la alta precisión obtenida para todas las categorías. Esta precisión podría incluso mejorarse, corrigiendo las inexactitudes y ambigüedades del conjunto de datos de prueba utilizado.

Estos buenos resultados se pueden ver claramente si se comparan con los obtenidos cuando el modelo utilizaba 7 categorías. Esta comparativa se puede observar en la Figura 9, que muestra la evolución de la precisión y de la función error si se comparan los modelos CCN número 3 (con 7 categorías) y CCN número 4 (con 4 categorías). Las líneas de tendencia son similares en ambos modelos. Sin embargo, el valor de precisión y la velocidad de convergencia son mucho mejores en segundo caso.

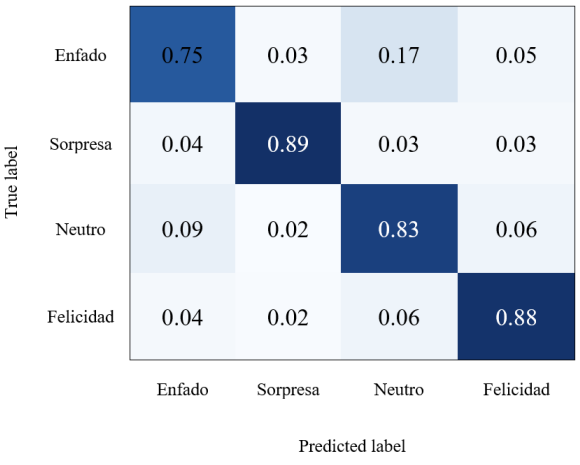


Fig. 8. Matriz de confusión del tercer modelo CCN con cuatro categorías

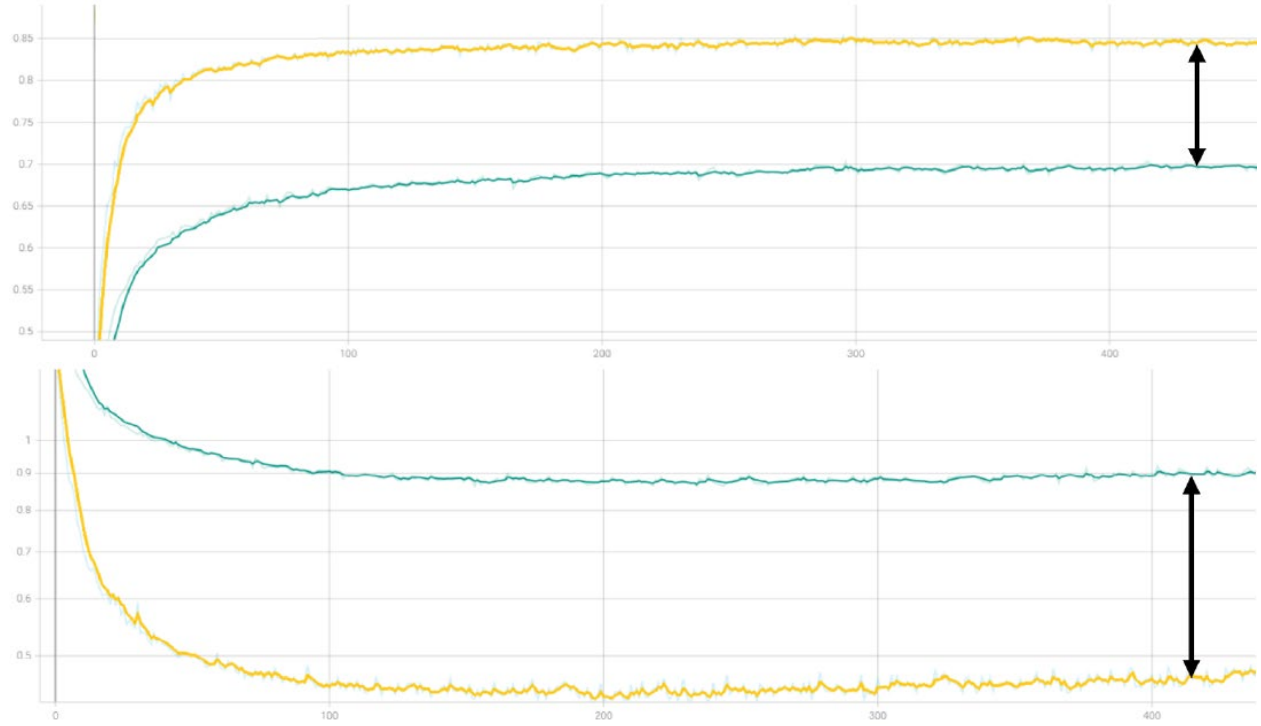


Fig. 9. Comparativa entre modelo CCN #3 (verde) y modelo CCN #4 con 4 categorías (naranja) a) Evolución de la precisión y b) Evolución de la función de error según la época (eje X)

V. Pruebas de Concepto

Para experimentar con el prototipo se hicieron unas pruebas de concepto en un entorno controlado utilizando como base el videojuego *Need for Speed Payback*, tal y como se puede observar en la Figura 10. El usuario tenía que conducir en el videojuego utilizando un volante y unos pedales. Al mismo tiempo, el prototipo analizaba las expresiones de su cara. Durante las pruebas realizadas, se pudo comprobar que el sistema es capaz de identificar correctamente emociones en el usuario. Para ello, los participantes en las pruebas debían forzar sus emociones de manera que se pudiese constatar

Fig. 10. Pruebas de concepto utilizando un videojuego

En la Figura 10 se muestra un instante de las pruebas, con el videojuego proyectándose en una pantalla y el sistema desarrollado posicionado en la parte delantera inferior. Tal y como se puede observar en la figura, el panel integrado en el prototipo muestra la emoción 5: "Sorpresa", que se ha estimado a partir de la observación de la cara del usuario.

VI. Conclusiones y Trabajos Futuros

En este trabajo se presenta un prototipo de un sistema de detección de emociones diseñado para un entorno de conducción segura. El prototipo es capaz de inferir información sobre el estado anímico de los conductores a partir de imágenes que se capturan durante la conducción. El propósito final de este prototipo es el de ser incorporado a un ADAS que integre información del vehículo, del entorno y del conductor. Mediante la combinación de datos recogidos de diferentes fuentes y utilizando la información del estado del conductor que es capaz de inferir el prototipo, será posible proporcionar recomendaciones orientadas a mejorar la seguridad en la conducción que se ajusten totalmente a las necesidades del usuario.

El sistema implementado utiliza técnicas de aprendizaje automático para poder determinar el estado anímico del conductor a partir de su expresión facial. El modelo subyacente se entrena fuera de línea y se puede cargar en un sistema embarcado, constituyendo un auténtico sistema de *edge computing* y garantizando su funcionamiento con independencia del grado de conectividad disponible en cada momento.

Se han implementado diferentes modelos que, mediante el entrenamiento, alcanzan unos valores notables de precisión. Durante la experimentación con estos modelos, se ha observado que los modelos generalistas que han sido pre-entrenados no proporcionan buenos resultados. Emplean demasiadas capas convolucionales que no consiguen una adecuada extracción de características. Por otro lado, con los modelos ad hoc se han obtenido buenos resultados. Sin embargo, los datos de

entrenamiento han supuesto un problema añadido. En primer lugar, la homogeneidad de las imágenes y la ambigüedad en algunos casos, han dificultado el proceso de inferencia. Prescindiendo de los datos más ambiguos y enriqueciéndolos mediante un pre-procesado de imágenes se han mejorado los resultados en gran medida. En segundo lugar, el hecho de que cada imagen estuviese asociada a una única emoción, también ha supuesto un problema adicional.

El sistema se basa en un modelo que se entrena fuera de línea. En un entorno de producción real, el modelo podría actualizarse de forma automática desde un sistema centralizado o en la nube, con procesos sucesivos de aprendizaje que permitiesen mejorar su precisión. Esto podría considerarse una de las líneas futuras de trabajo, junto a la ya comentada de disponer de un ADAS totalmente integrado que utilice los datos proporcionados del sistema junto con información recogida de otras fuentes. Adicionalmente, en su implementación actual el sistema reporta las emociones detectadas de forma inmediata, lo que puede dar lugar a falsos positivos. La precisión mejorará con la implementación de métricas combinadas en el ADAS y con los algoritmos a implementar de cara a ofrecer las recomendaciones al conductor. Finalmente, sería necesario contrastar el funcionamiento del sistema mediante pruebas de conducción en un entorno real.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por el Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación 2013-2016, a través del proyecto TIN2017-82928-R.

REFERENCIAS

[1] J. R. Treat, N. S. Tumbas, S. T. McDonald, D. Shinar, R. D. Hume y R. E. Mayerm, "Tri-level study of the causes of traffic accidents: Interim report I, volume I research findings" NHTSA, USA, Informe técnico DOT HS-805 085, 1979.

[2] D. L. Hendricks, J. C. Fell y M. Freedman, "The relative frequency of unsafe driving acts in serious traffic crashes", NHTSA, USA, Informe técnico DTNH22-94-C-05020, 1999.

[3] "Zen driving, conducción zen: emociones, decisiones y conducción", UNESPA, 2012.

[4] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang y Y. Bengio. "Challenges in representation learning: A report on three machine learning contests", en *Neural Networks*, vol. 64, 2015, pp. 59-63.

[5] S. Paiva, X.G. Pañeda, V. Corcoba, R. García, P. Morán, L. Pozueco, M. Valdés, C. del Camino, "User Preferences in the Design of Advanced Driver Assistance Systems", en *Sustainability*, vol. 13, nº 7, 2021.

[6] A. Williamson y T. Chamberlain, "Review of on-road driver fatigue monitoring devices", 2005.

[7] L. Bretzner y M. Krantz, "Towards low-cost systems for measuring visual cues of driver fatigue and inattention in automotive applications" en actas del *IEEE International Conference on Vehicular Electronics and Safety*, China, 2005, pp. 161–164.

[8] Q. Wang, J. Yang, M. Ren y Y. Zheng, "Driver Fatigue Detection: A Survey", en actas del *6th World Congress on Intelligent Control and Automation*, 2006, pp. 8587-8591.

[9] V. Balasubramanian y K. Adalarasu, "EMG-based analysis of change in muscle activity during simulated driving", en *Journal of Bodywork and Movement Therapies*, vol. 11, nº 2, 2007, pp. 151-158.

[10] S.J. Jung, H.S. Shin, y W.Y. Chung, "Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering Wheel", en *IET Intelligent Transportation Systems*, vol. 8, nº 1, pp. 43–50, 2014.

[11] S. Shaily, S. Krishnan, S. Natarajan, et al. "Smart driver monitoring system", en *Multimedia Tools and Applications*, vol. 80, pp. 25633–25648, 2021

[12] N. Moslemi, M. Soryani, R. Azmi, "Computer vision-based recognition of driver distraction: A review", en *Concurrency and Computation Practice and Experience*, 2021, https://doi.org/10.1002/cpe.6475

[13] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," en *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.2981446.

[14] K. Simonyan y A. Zisserman, "Very deep convolutional networks for large-scale image recognition", en actas del *3rd International Conference on Learning Representations*, USA, 2015.

[15] S. Zhao, H. Cai, H. Liu, J. Zhang, and S. Chen, "Feature selection mechanism in CNNs for facial expression recognition", en actas del *29th British Machine Vision Conference*, Reino Unido, 2018

[16] P. Viola y M. Jones, "Rapid object detection using a boosted cascade of simple features", en actas del *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, USA, 2001.

[17] J. M. Susskind, A. K. Anderson y G. E. Hinton, "The toronto face database," en Informe Técnico del Departamento de Informática de la Universidad de Toronto, Canadá, vol. 3, 2010.

[18] R. Gross, I. Matthews, J. Cohn, T. Kanade y S. Baker, "Multi-pie", en *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[19] C. F. Benitez-Quiroz, R. Srinivasan y A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," en actas del *IEEE International Conference on Computer Vision & Pattern Recognition* (CVPR), Las Vegas, USA, 2016.

[20] S. Li y W. Deng, "Reliable crowdsourcing and deep localitypreserving learning for unconstrained facial expression recognition", en *IEEE Transactions on Image Processing*, vol. 28, nº 1, pp. 356-370, 2019.

[21] A. Mollahosseini, B. Hasani y M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild", en *IEEE Transactions on Affective Computing*, vol. 10, nº 1, pp. 18–31, 2019.

[22] Z. Zhang, P. Luo, C. L. Chen y X. Tang, "From facial expression recognition to interpersonal relation prediction", en *International Journal of Computer Vision*, vol. 126, pp. 550–569, 2018

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

OPPNets and rural areas: an opportunistic solution for remote communications

Manuel Jesús-Azabal, Juan Luís Herrera, Sergio Laso y Jaime Galán-Jiménez
Departamento de Sistemas Informáticos y Telemáticos
Universidad de Extremadura
Avda. de la Universidad, S/N, Cáceres, España
{manuel, jlherrerag, slasom, jaime}@unex.es

***Resumen*— Many rural areas along Spain do not have access to the Internet. The lack of communication infrastructures in these zones hinders the deployment of technological health solutions, affecting the quality of life of the population, which use to be people over 65 who live alone. Also, these circumstances impact on local businesses which are widely related to the agricultural and livestock industry. Considering this situation, we propose a solution based on an opportunistic network algorithm which enables the deployment of technological communication solutions for both elderly healthcare and livestock industrial activities. The algorithm bases the transmission on the interest of the nodes and on its collaboration. Thus, the solution has been evaluated considering several simulations under multiple conditions, comparing the delivery probability, latency and overhead outcomes with other well-known opportunistic routing algorithms, outperforming them. As such, the proposed approach provides a reliable mechanism for the data transmission in isolated rural scenarios.**

***Palabras Clave*- Opportunistic Networks, Internet of Things, Routing Algorithms, Elderly, Healthcare, Offline.**

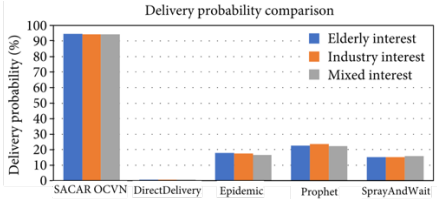
I. SUMMARY

Many isolated rural areas do not have an Internet connection. These limitations are often induced by the remote situation and the geographical issues, as well as the lack of interest from telecommunication companies. This aspect has an impact on the population, which is usually over 65 years old and requires services and health attention, especially when these elderly people live alone. This context can be improved with digital healthcare, which is an active work line [1]. Nevertheless, the lack of an Internet infrastructure prevents the deployment of many of these solutions. Moreover, the technological gap also affects the local industry and livestock, which are not generally able to implement solutions to monitor and improve performance.

In order to face these challenges, we propose a routing algorithm based on opportunistic networks which transfers elderly presence information and industry performance data to gateway points. For this, the solution bases the routing on the interests of the nodes, communicating the source and the destination with intermediate entities such as cars, pedestrians and throwboxes which are subscribed to concrete interests. As a result, the traffic is transmitted along the nodes of the network which are interested on the topic of the data. For this work, two topics are considered: presence information of the elders to detect potential emergencies; and performance data about the industries.

The algorithm has been simulated on an realistic isolated region of Spain. For this, variables such as delivery probability, latency, overhead and hops have been considered. The results obtained in the executions are positive and guarantee a high rate of successful message delivery under different conditions. Furthermore, the outcomes of the algorithm are also compared with other well-known opportunistic routing solutions. As a result, the proposed solution quadruples the delivery probability of Prophet, which presents the best results among the benchmark solutions; and greatly reduces the overhead regarding other solutions such as Epidemic or Prophet.

In conclusion, the proposed scheme brings a solution to the technological gap in isolated rural areas, enabling the monitoring of elders' activity and providing a reliable communication system to improve the performance of livestock industries.



REFERENCIAS

[1] Sekhon, H., Sekhon, K., Launay, C., Afililo, M., Innocente, N., Vahia, I., Rej, S., Beauchet, O.: Telemedicine and the rural dementia population: a systematic review. Maturitas 143, 105–114 (2020)

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

GirolA: Una pasarela de servicios web

Antonio Delgado, Antonio Estepa.
Departamento de Ingeniería Telemática,
Universidad de Sevilla
C/ Camino de los Descubrimientos s/n
{aldelgado,aestepa}@us.es.

La integración entre múltiples aplicaciones con varios proveedores de servicios web puede resultar un problema complejo en organizaciones de mediano y gran tamaño donde el parque de aplicaciones y sistemas destino con los que es necesario integrarse se multiplica. Además, la variabilidad de mensajes y tecnologías implicadas en este tipo de entornos no hacen más que incrementar la complejidad de la situación.

En este artículo se presenta GirolA, una pasarela de servicios web diseñada con el objetivo de facilitar la integración de aplicaciones en entornos como el mencionado y que lleva siendo empleada con éxito durante más de un año. En dicho periodo se han integrado 19 servicios y se han realizado más de 2.7M de invocaciones a servicios web por lo que lo consideramos un caso de éxito.

***Palabras Clave*- pasarela de servicios, integración**

I. INTRODUCCIÓN

La integración de los sistemas de información es una de las necesidades básicas en las grandes organizaciones. Hoy es habitual que una aplicación se comunique con otras para consultar su información o solicitarles que realicen determinadas acciones. Afortunadamente, disponemos de tecnologías fuertemente asentadas que proporcionan los pilares para desarrollar las soluciones de integración que pudieran ser necesarias: SOAP, REST, UDDI, WSDL, etc. se han convertido en estándares de facto para esta labor [1, 2]. Sin embargo, satisfacer las necesidades de integración en una organización de gran tamaño, en especial del sector público, puede llegar a ser muy complejo. En casos extremos, un organismo puede manejar cientos de aplicaciones y cada una de ellas puede necesitar integrarse con varios sistemas destino que, a su vez, pueden emplear distintas tecnologías de integración obligando a nuestras aplicaciones a contemplar cada una de ellas. Incluso cuando todos los sistemas destino con los que tenga que interactuar una de nuestras aplicaciones empleen una misma tecnología de servicios web existirán distintos mensajes a intercambiar con ellos, con una gran variabilidad en su estructura y sintaxis. Todo ello conlleva un gran esfuerzo de desarrollo que debe mantenerse prácticamente en cada sistema de la organización. Por último, tampoco es extraño que en organizaciones de gran tamaño exista un buen número de aplicaciones heredadas con un mantenimiento muy complicado o incluso imposible de llevarse a cabo en algunos casos.

En este contexto podemos tomar algunas medidas para reducir la presión sobre los equipos de desarrollo. En algunos casos se puede optar por atacar la problemática de la integración desarrollando bibliotecas que abstraigan a nuestras aplicaciones de la tecnología y mensajes concretos a intercambiar con el sistema destino. Es una buena solución, pero no carente de algunos inconvenientes. En primer lugar, estas bibliotecas deben mantenerse actualizadas lo que deriva en evolutivos que deberán ser implementados en nuestras aplicaciones. Por ejemplo, si cierto campo que era opcional en una llamada pasa a ser obligatorio, la biblioteca deberá ser actualizada y también las aplicaciones que hagan uso de esa llamada. De hecho, no hace falta ir a un cambio funcional para ilustrar este problema. Cualquier nueva compilación de la biblioteca obligará a recompilar todas las aplicaciones que dependan de ella. Por lo tanto, las bibliotecas, aunque amortiguan el esfuerzo necesario para mantener nuestras aplicaciones integradas con los sistemas destino, no lo eliminan por completo. Desplazan el esfuerzo del desarrollo a un único punto, pero se mantiene un esfuerzo de desarrollo, aunque sea mínimo, en nuestras aplicaciones. En segundo lugar, no es infrecuente que existan aplicaciones en las que no se pueda hacer uso de las bibliotecas para resolver sus necesidades de integración debido a diferencias de versiones, uso de tecnologías no compatibles u otros problemas [3]. Por último, hay que destacar que esta solución será prácticamente imposible de aplicar en sistemas heredados ya que en muchos casos carecerán de un mantenimiento adecuado.

En este artículo se propone una solución alternativa para solventar algunos de los problemas comentados con anterioridad, GirolA, una pasarela o *gateway* de servicios

web [4]. Un sistema de este tipo permite desacoplar a las aplicaciones de las bibliotecas de integración. Si además el *gateway* admite la integración con los sistemas destino a través mecanismos alternativos (como el uso de ficheros) puede permitir resolver los problemas de integración de las aplicaciones heredadas. Además, veremos que podemos dotar a una plataforma de este tipo de muchas funcionalidades interesantes que pueden elevar el nivel de integración de nuestra organización de manera importante.

En las siguientes secciones de este artículo se describe de forma general el sistema desarrollado para continuar profundizando en su conocimiento a través del modelo de dominio, el de presentación (interfaz de usuario), el modelo de procesos y su interfaz de programación de aplicaciones (API). Una vez presentado el sistema se relata nuestra experiencia práctica con él durante más de un año de uso. Por último, cerrando el documento, se plantean algunas conclusiones obtenidas de nuestra experiencia particular.

II. Estado del arte

Un *gateway* o pasarela de servicios web es un artefacto software que actúa de intermediario en la comunicación entre un consumidor y un proveedor de servicios web. Las pasarelas de servicios web proporcionan un único punto de control, acceso y validación de solicitudes de servicios web, y permiten controlar qué servicios están disponibles para los distintos consumidores de los servicios web [5].

En la literatura podemos encontrar *gateways* de servicios web con distintos propósitos:

- Traducción de protocolos de servicios web o tecnologías de distribución de procesamiento. Probablemente la forma más sencilla de convertir un servicio SOAP a un servicio REST o viceversa sea usar una pasarela de servicios web (en estos casos también conocidos como un proxy de servicio). Estos sistemas realizan las traducciones entre los mundos SOAP y REST de las peticiones y las repuestas pertinentes. Un producto que realiza esta función es *Membrane Service Proxy* [6]. También existen publicaciones en las que se realizan propuestas en esta línea [7]. En algunos casos incluso se han propuesto sistemas más complejos que actúan como pasarela entre tecnologías distintas para procesamiento distribuido [8].
- Seguridad. La eclosión de los servicios web ha incrementado la exposición de recursos críticos generando un mayor esfuerzo para la securización de los sistemas que puede ser mitigado con el uso de *gateways* para servicios web [9]. Otras propuestas relativas a la seguridad se han centrado en evitar ataques. Por ejemplo, en [10] se presenta *Checkway*, un *gateway* para proteger servicios web de ataques de denegación de servicio. Por último, en [11] se propone el uso de una pasarela que controla el acceso a un nivel muy fino en las invocaciones a los servicios web.
- Un tercer grupo de aportaciones está orientado a resolver el problema de la integración con múltiples orígenes de datos manteniendo una consistencia en la información obtenida con el objeto de realizar posteriores análisis o estudios. En esta línea *TogoWS* [12] ofrece una solución en el dominio específico de la biotecnología.

Como hemos visto, la idea de emplear pasarelas de servicios web no es nueva y su uso ha sido propuesto con fines diversos a lo largo del tiempo. Sin embargo, no existen propuestas que aconsejen su empleo como solución al problema que ocasiona la integración en organizaciones con un alto número de aplicaciones. Tampoco hay estudios dirigidos al objetivo de reducir la carga de desarrollo que implica el que una aplicación tenga que integrarse con distintos sistemas cada uno con tecnologías y mensajes diferentes. Tampoco hay propuestas enfocadas en tratar de ofrecer una solución en este sentido para aplicaciones heredadas.

III. Descripción general del sistema

A grandes rasgos, nuestro objetivo es desarrollar un *gateway* de servicios web que simplifique la integración de las aplicaciones de una organización con los distintos sistemas destino proveedores de servicios web. Los principales requisitos que para nosotros debe cumplir el sistema se enumeran en la Tabla I.

Tabla I
Requisitos principales del sistema

Requisito	Descripción
1	El sistema debe permitir la integración con sistemas destino de distintas tecnologías (SOAP, REST, etc.), tipos (síncronos o asíncronos) y que empleen distintos métodos para la autenticación, (usuario + clave, certificados, etc.)
2	La pasarela debe interactuar con las aplicaciones de la organización, pero también con usuarios de distinto perfil
3	Las peticiones recibidas por el *gateway* (también denominadas lotes o trabajos) estarán formadas por un grupo de solicitudes individuales. Una petición estará compuesta por lo tanto al menos por una solicitud individual
4	El tratamiento de las peticiones recibidas por la pasarela será asíncrono
5	Los trabajos o lotes podrán comunicarse a la pasarela por diversos medios: ficheros, servicio web, base de datos, etc.
6	El sistema no debe limitarse a ser un mero intermediario en las peticiones que se le realicen para integrarse con algún sistema externo. Debe tratar la información obtenida de esos sistemas externos simplificando la respuesta para las aplicaciones de la organización
7	La sintaxis de las peticiones y sus respuestas deben mantener una estructura homogénea y ser independientes del servicio final invocado
8	El sistema debe almacenar información sobre todas las transacciones realizadas de forma que puedan utilizarse para auditorías u otros usos

En la figura 1, se muestra un diagrama de contexto del sistema. En esta figura se puede apreciar que las aplicaciones de la organización (APP1..APP3) disponen de tres vías para realizar peticiones de integración con sistemas remotos: a través de ficheros CSV empleando un servidor SFTP, empleando una API de servicios web o utilizando un conjunto de funciones definidas en la base de datos del sistema (requisitos 2 y 5). También vemos que los usuarios pueden interactuar con el sistema mediante una interfaz de usuario que les permite generar trabajos utilizando también ficheros en formato CSV (requisitos 2 y 5). La pasarela, es capaz de interactuar con sistemas externos de distinta naturaleza (requisito 1). En la figura aparecen GIRO (el sistema de Gestión Integral de Recursos Organizativos de la Junta de Andalucía), SCSP (Supresión de Certificados en Soporte Papel) [13] y otros como podrían ser los servicios web de la AEAT [14], pero la plataforma permite la integración sencilla con cualquier proveedor de servicios web. En el diagrama también se quiere poner de relieve que el *gateway* no es un mero "intermediario" de peticiones si no que tiene cierta capacidad de proceso y que es capaz de tratar las respuestas recibidas, traduciéndolas o simplificándolas para el solicitante y siendo capaz de generar hasta tres salidas diferentes para un lote de peticiones (requisito 6). Por último, también se quiere destacar que el sistema almacena en su base de datos información de todas las transacciones realizadas, así como trazas (log) de las peticiones efectuadas y las respuestas obtenidas de los sistemas destino (requisito 8).

De manera simplificada podemos decir que GirolA recibe lotes de trabajos por distintos medios y de distintos solicitantes: aplicaciones o humanos. Esos trabajos son encolados y se van ejecutando efectuando las invocaciones a los servicios web pertinentes en cada caso. La pasarela se encarga de componer las llamadas a los servicios web en la forma adecuada, autenticarse en el sistema destino en la manera en que sea necesario y obtener la respuesta del mismo ya sea de manera síncrona o asíncrona. Una vez obtenida la respuesta el sistema la almacena y tiene la capacidad de tratarla para ofrecer a los solicitantes una información más refinada. La respuesta puede ser tratada de tres formas distintas teniendo así la posibilidad de generar hasta tres resultados diferentes para un trabajo concreto.

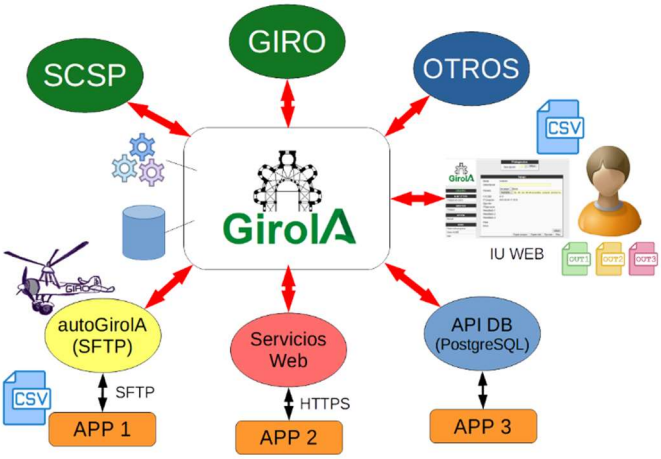


Fig. 1. Diagrama de contexto del sistema

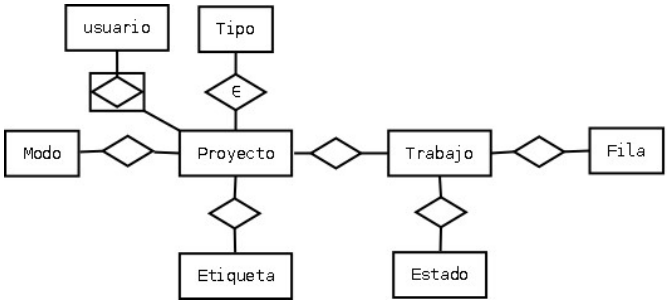


Fig. 2. Diagrama Entidad-Relación

IV. Modelo de dominio

El modelo de dominio de nuestro sistema (Fig. 2) recoge principalmente información sobre proyectos de integración, los trabajos y su contenido. A continuación, se describen las entidades más relevantes del modelo.

Proyecto. Recoge los datos de lo que en GirolA se denomina un proyecto de integración. Un proyecto de integración define una forma determinada de interactuar con alguno de los servicios web soportados por GirolA. Los atributos principales de esta entidad son el servicio con el que se desea trabajar, el entorno al que atacará el proyecto (pruebas, preproducción, producción), usuario y clave de acceso, etc. También son atributos de esta entidad una serie de ficheros que definen cómo se ha de generar la llamada al servicio web en cuestión y el tratamiento que se desea realizar de la respuesta obtenida. De todos ellos podemos destacar los siguientes:

- `callh`, `callb` y `callt`: cabecera, cuerpo y cola de la llamada. Se trata de ficheros de texto que actúan a modo de plantilla para componer el XML o JSON que se empleará para realizar la llamada al servicio web.
- `dbproc`, `proc1`, `proc2` y `proc3`: especifican una transformación xslt o un filtro jq [15] para realizar tratamientos de la respuesta obtenida. El primero de estos ficheros es obligatorio y se encarga de extraer de la respuesta la información necesaria para alimentar la base de datos de GirolA. Los otros tres son opcionales y posibilitan especificar hasta tres tratamientos alternativos de la respuesta para generar distintos ficheros de salida.

Tipo. Los proyectos de integración se clasifican según su tipo. Esta entidad sólo dispone de una clave primaria secuencial y una descripción.

Trabajo. Asociados a un proyecto de integración se pueden generar infinidad de trabajos. Un trabajo tiene una descripción que debe ser única, un usuario creador, un usuario ejecutor, fechas de creación y ejecución, etc.

Estado. Cada trabajo tiene asociado un estado que tiene relación con la ejecución del mismo. Así un trabajo

puede estar pendiente de ejecución, pendiente de respuesta asíncrona, finalizado con éxito, finalizado con error, etc.

Fila. Esta entidad define cada uno de datos individuales que componen un trabajo o lote. Para cada fila se pueden informar hasta 20 columnas que recogen datos que pueden ser empleados tanto en la invocación del servicio web como en la generación de resultados.

Etiqueta. Una etiqueta sirve para definir un comportamiento en cuanto a la ejecución y planificación de los trabajos asociados a un proyecto e incluso si fuera necesario establecer prioridades distintas en el tratamiento de los mismos. Los trabajos no etiquetados siguen la política general: se encolan por orden de llegada y posteriormente se van extrayendo de la cola y se ejecutan en paralelo hasta un máximo de seis procesos. Una etiqueta puede determinar el uso de colas particulares o planificaciones distintas para la ejecución ciertos trabajos. Actualmente están definidas en el sistema las siguientes etiquetas:

- `1_POR_MINUTO`: Esta etiqueta define una cola en la que los trabajos se extraen cada minuto y se ejecutan de uno en uno. Está destinada trabajos que emplean servicios web muy pesados en los que una invocación puede llegar a ocupar 1.5MB
- `3_EN_PARALELO`: Esta etiqueta define una cola en la que los trabajos son ejecutados como máximo en grupos de tres.
- `NOCTURNO`: Define una planificación similar a la que se emplea por defecto pero en la que la ejecución de los trabajos se permite únicamente entre las 00:00 y las 06:00.

Modo. El modo define el comportamiento del proyecto de integración. Actualmente están definidos los siguientes modos:

- LITIS: Llamada Individual con Tratamiento Individual (servicio destino Síncrono)
- LCTIS: Llamada Colectiva con Tratamiento Individual (servicio destino Síncrono)
- LCTCS: Llamada Colectiva con Tratamiento Común (servicio destino Síncrono)
- LCTUS: Llamada Colectiva con Tratamiento Único (servicio destino Síncrono)
- LCTIA: Llamada Colectiva con Tratamiento Individual (servicio destino Asíncrono)
- LCTCA: Llamada Colectiva con Tratamiento Común (servicio destino Asíncrono)
- LCTUA: Llamada Colectiva con Tratamiento Único (servicio destino Asíncrono)

La llamada individual (LI) indica que por cada fila del trabajo o lote se debe realizar una invocación al servicio web especificado en el proyecto de integración. Por el contrario, una llamada colectiva (LC) implica que hay que componer una única invocación que contenga información sobre cada una de las filas que componen el lote.

Respecto a los tratamientos de la respuesta proporcionada por el servicio web se distinguen tres tipos:

1. Tratamiento individual (TI): la respuesta obtenida se trata individualmente para cada fila que compone el lote.
2. Tratamiento común (TC): la respuesta obtenida es común para todas las filas que componen el lote y por lo tanto los valores respuesta para cada fila del lote son idénticos
3. Tratamiento único (TU): la respuesta obtenida es común para todas las filas del lote, sin embargo, no se va a generar un resultado individualizado, el resultado será único para el lote.

El último carácter indica el modo de ejecución del servicio web destino que puede ser síncrono (S) o asíncrono (A).

V. Modelo de presentación

La interfaz de usuario del sistema ha sido desarrollada empleando WAINE [16], un entorno de desarrollo de interfaces de usuario basado en modelos. En ella, se definen cuatro roles: administrador, usuario avanzado, usuario regular y consulta. En la Tabla II se especifican las unidades de interacción accesibles a cada rol.

En los siguientes apartados se describen las unidades de interacción de mayor relevancia (destacadas en negrita en la Tabla II)

A. Gestión de proyectos de integración

Esta interfaz de usuario (Fig. 3) permite a los administradores gestionar proyectos de integración, así como asignar la visibilidad sobre el proyecto a los distintos usuarios (zona inferior de la unidad de interacción). En la pestaña información se cumplimentan los campos principales de la entidad. En las pestañas *Ficheros Obligatorios*, *Ficheros Opcionales*, *Procesamiento Extra 1*, *Procesamiento Extra 2* y *Procesamiento Asíncrono* se adjuntan los ficheros que fueran necesarios para definir las invocaciones a los servicios web y especificar los distintos tratamientos de la respuesta.

Tabla II
Rol-funcionalidad

	Admin	U.avanzado	Usuario	Consulta
Gest. Proyectos de integración	✓			
Gest. Categorías	✓			
Gest. Etiquetas	✓			
Gest. Trabajos	✓	✓	✓	
Trabajos en curso	✓	✓	✓	
Cons. Últimos trabajos realizados	✓	✓		
Cons. Trabajos por proyecto	✓	✓		
Búsqueda avanzada	✓	✓		
Búsqueda resultado	✓	✓		
Estad. Trabajos por categoría	✓			
Estad. Trabajos anuales	✓			
Gestión de usuarios	✓			
Búsqueda de resultados reducida			✓	✓

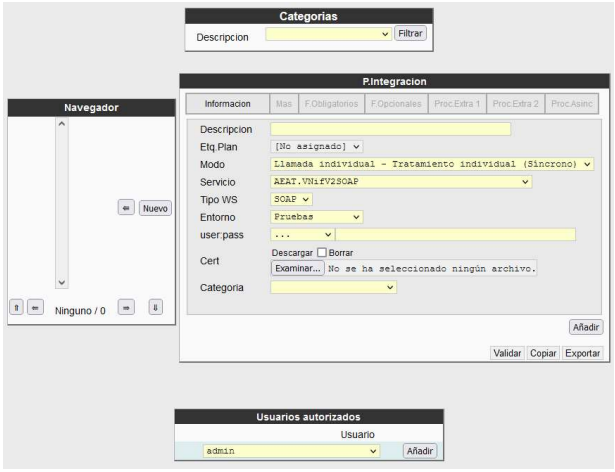


Fig. 3. Gestión de proyectos de integración

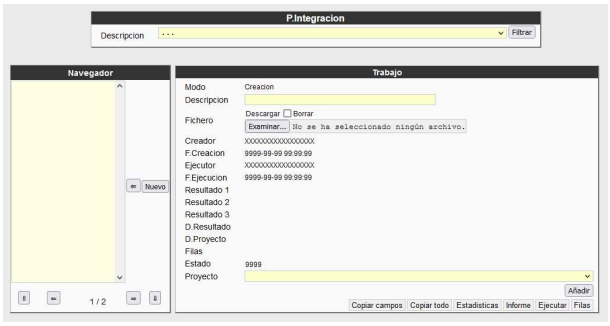


Fig. 4. Gestión de trabajos

B. *Gestión de trabajos*

Esta unidad de interacción (Fig. 4) permite el alta, modificación y eliminación de trabajos sobre los proyectos de integración a los que el usuario tiene acceso. También permite realizar diversas acciones como copiar el trabajo, extraer estadísticas de los resultados, generar un informe sobre la ejecución del mismo, ejecutar el trabajo y por último ver el detalle de las filas que lo componen.

C. *Trabajos en curso*

Presenta para cada trabajo en ejecución una serie de campos que dan una idea de su velocidad de ejecución y su evolución esperada (ver Fig.5). Entre ellos aparecen el usuario ejecutor, la fecha de inicio, el número de peticiones por minuto (ppm), la duración estimada del trabajo (de) y la hora estimada de fin (hef). También permite ver el detalle de las filas que componen el trabajo, obtener estadísticas del mismo y acceder a la carpeta donde se almacenan las peticiones y respuestas de cada invocación al proveedor de servicios web realizada en el proyecto.

Fig. 5 Trabajos en curso

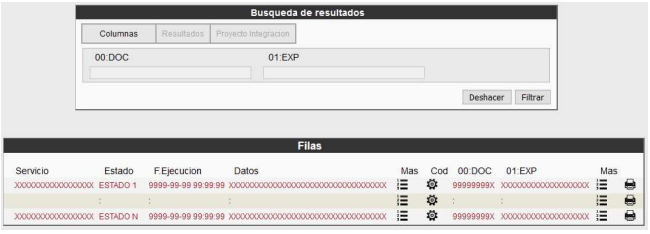


Fig. 6. Búsqueda de resultados reducida

D. *Búsqueda de resultados reducida*

Para finalizar esta sección presentamos la interfaz de usuario "búsqueda de resultados reducida" (Fig. 6). Esta unidad de interacción permite localizar resultados en función de la información que se aportó para la invocación al servicio.

VI. PROCESOS

En esta sección se presentan los procesos más relevantes del sistema. En la figura 7 podemos ver un diagrama que ilustra los distintos procesos y las dependencias existentes entre ellos, así como las relaciones que guardan con otros artefactos. A continuación, se describe brevemente cada proceso.

girola.action: Es la acción lanzada por el botón Ejecutar de la interfaz de usuario gestión de trabajos (ver Fig. 3). Este proceso toma el fichero CSV asociado al trabajo y lanza una ejecución manual del mismo invocando al proceso girola.man

girola.auto: Este proceso da soporte a lo que se ha denominado autogirola. Realiza un proceso similar al que realiza girola.action, salvo que el trabajo manual es lanzado por la creación de una carpeta en un directorio determinado que contiene una relación de ficheros que deben ser procesados por la pasarela.

girola.man: Se encarga de la ejecución de trabajos que vienen definidos por un fichero CSV. Este proceso da de alta el trabajo con sus correspondientes filas en la base de datos del sistema e inmediatamente invoca a girola.exec para que realice la ejecución del mismo.

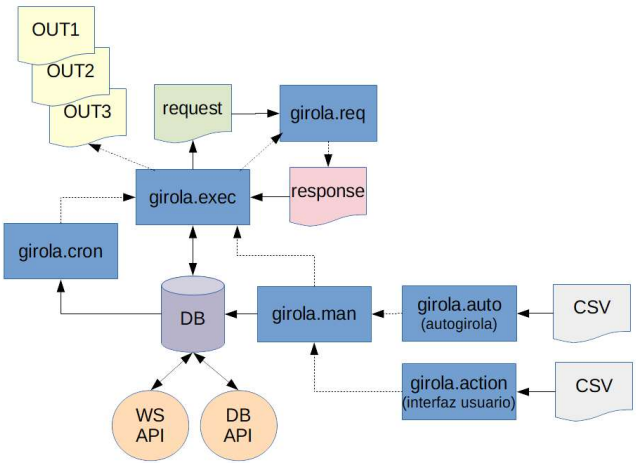


Fig. 7. Representación de los procesos

girola.cron: Realiza una tarea similar a la de girola.man pero para aquellos trabajos que han sido generados a través de la API de la base de datos o de la API de servicios web de GirolA. Empleando este proceso en combinación con las etiquetas de los proyectos de integración (ver sección IV) se pueden definir distintas planificaciones de ejecución para los trabajos. El funcionamiento básico de girola.cron consiste en tomar el trabajo más antiguo aún no ejecutado para una etiqueta dada y lanzar su ejecución invocando a girola.exec según una planificación establecida en el fichero crontab [17] e identificada por la etiqueta asignada al proyecto.

girola.exec: Es el núcleo de la ejecución de procesos. Este proceso toma cada fila del trabajo a ejecutar, monta las llamadas SOAP o REST, realiza la llamada al servicio web pertinente invocando a girola.req, trata las respuestas para generar los ficheros de salida necesarios y almacena los resultados en la base de datos de GirolA. Durante el proceso genera un log del trabajo y almacena evidencias de la ejecución en el sistema de ficheros.

girola.req: es el encargado de realizar llamadas a los distintos servicios web, abstrayendo a otros procesos de asuntos propios de la llamada como pueden ser el *host* involucrado, el *endpoint* a utilizar, la autenticación o firma de la llamada si fuera necesaria, etc.

VII. INTERFAZ DE PROGRAMACIÓN DE APLICACIONES

No se podría finalizar esta exposición del sistema sin presentar el conjunto de funciones que permiten a las aplicaciones de la organización integrarse con GirolA. De manera abstracta la API tanto de servicios web (actualmente SOAP) como la soportada por la base de datos (ver Fig. 1) está compuesta por las siguientes funciones:

`1. ID addLDRow(IN cuser,IN proj,IN lddescr,IN col00,IN,..,IN col19)`. Añade una fila con las columnas proporcionadas (col00 .. col19) a un lote o carga identificado por lddescr. En la llamada se informan también el usuario o aplicación que realiza la inserción (cuser) y el proyecto de integración que se debe emplear (proj). Durante la primera inserción de una fila en el lote, se crea el mismo. Se pueden seguir añadiendo filas al lote mientras no se le haya asignado tipo al lote utilizando la función `setLoadType`. Si la función tuvo éxito devuelve el identificador de la fila insertada. A continuación, se muestra un ejemplo de la invocación a esta función empleando el API en base de datos y el API SOAP.

API en base de datos

Llamada

```
select addLDRow('AplicacionX', 65, 'LOTE-1',
'99999999K', 'ES0123456789012345678901');
```

Respuesta

```
addldrow
----------
         6
```

API SOAP

Llamada

```
<soapenv:Envelope xmlns:soapenv=
"http://schemas.xmlsoap.org/soap/envelope"
xmlns:girola="http://juntadeandalucia.es/GIROLA">
<soapenv:Header/>
   <soapenv:Body>
      <ADDLDROWREQ>
         <HEAD>
            <APP>AplicacionX</APP>
         </HEAD>
        <PARAMS>
         <PROJ>65</PROJ>
         <DESCR>LOTE-1</DESCR>
         <COLS>
           <ITEM>99999999K</ITEM>
           <ITEM>ES0123456789012345678901</ITEM>
         </COLS>
        </PARAMS>
      </ADDLDROWREQ>
   </soapenv:Body>
</soapenv:Envelope>
```

Respuesta

```
<?xml version="1.0"?>
<SOAP:Envelope xmlns:SOAP=
    "http://schemas.xmlsoap.org/soap/envelope/">
   <SOAP:Header/>
   <SOAP:Body>
     <ADDLDROWRESP>
     <HEAD>
      <ERROR>000</ERROR>
      <ERRDESCR>Ejecucion sin errores</ERRDESCR>
     </HEAD>
     <RESULT>
       <ROWID>6</ROWID>
     </RESULT>
     </ADDLDROWRESP>
   </SOAP:Body>
</SOAP:Envelope>
```

`2. ID delLDRow(IN cuser, IN id)`. Elimina una fila de un lote de trabajo. La llamada recibe el usuario o aplicación que realiza la inserción (cuser) y el identificador de fila a eliminar (id). Esta eliminación sólo se puede realizar cuando el trabajo no ha sido ejecutado aún. Si tuvo éxito devuelve el identificador de la fila eliminada.

`3. getLDRowResult(IN cuser,IN id,OUT err,OUT stat,OUT ts,OUT data)`. Obtiene el conjunto de resultados básicos para la fila identificada por el identificador id.

`4. getLDRowFullResult(IN cuser,IN id,OUT err,OUT stat,OUT ts,OUT data,OUT r00, .. ,OUT r19)`. Obtiene todos los resultados para una fila, los básicos (los mismos que se obtienen con una llamada a `getLDRowResult`) y otros que pueden haber sido definidos por el proyecto de integración. El proyecto de integración puede definir hasta 20 resultados para una llamada.

`5. Boolean isLoadFinished(IN cuser, IN d)`. Dado un usuario creador y un lote la función devuelve verdadero si el trabajo está finalizado y falso en caso contrario.

`6. ID setLoadType(IN cuser, IN d, IN t)`. Establece el tipo de trabajo para un lote determinado, M: Manual, A: Automático. Una vez asignado no se pueden añadir más filas al trabajo. Si el estado se asigna a automático el sistema lanzará el trabajo de forma atendiendo a la planificación establecida.

Tabla III
LÍNEAS DE CÓDIGO

Fichero	Descripción	Líneas
Girola.asl	Código XML con la especificación de los modelos de usuario, diálogo y presentación	1805
Girola.sql	Código SQL correspondiente al modelo de dominio	624
Girola.action	Script correspondiente al proceso girola.action	63
Girola.auto	Script correspondiente al proceso girola.auto	84
Girola.cron	Script correspondiente al proceso girola.cron	54
Girola.exec	Script correspondiente al proceso girola.exec	719
Girola.man	Script correspondiente al proceso girola.man	61
Girola.req	Script correspondiente al proceso girola.req	114
	Total	3524

`7. syncLDRowResult(IN cuser,IN proj,IN lddescr,IN col00,..,IN col19,OUT err,OUT stat,OUT ts,OUT data,OUT r00,.., OUT r19)`. Realiza una llamada síncrona para una única fila con las columnas proporcionadas y devuelve los resultados. Esta función proporciona la única manera de realizar un trabajo síncrono en GirolA. Sólo está disponible en el API de servicios web (SOAP).

El API de la pasarela define un conjunto de mensajes que mantienen una estructura homogénea y que son independientes del servicio final invocado como se indicaba en el requisito 7 de la Tabla I. De hecho, en la mayoría de casos una aplicación sólo necesita emplear las funciones 1, 3, 5 y 6. Este conjunto mínimo es suficiente para dar de alta trabajos y obtener sus resultados básicos.

VIII. EXPERIENCIA CON GIROLA

GirolA comenzó a ser construido en abril de 2020 y actualmente su desarrollo está prácticamente finalizado. El proyecto tiene un total de 3524 líneas de código distribuidas como se muestra en la Tabla III.

Las distintas funcionalidades soportadas se han ido incorporando al sistema atendiendo a las necesidades más prioritarias de la organización. En paralelo a su desarrollo se han ido definiendo diversos proyectos de integración y ejecutando trabajos sobre ellos. Actualmente hay definidos en el sistema 115 proyectos de integración sobre los que se han ejecutado unos 5.628 trabajos que contenían 2.784.732 filas (1.720.755 en 2020 y 1.063.977 hasta el momento de escribir este artículo en junio de 2.021). En la figura 8 se muestra un diagrama en el que se detalla la distribución de los trabajos y el número de filas a lo largo del tiempo.

Tabla IV
TOP 5 SERVICIOS WEB MÁS UTILIZADOS

Servicio web	Trabajos	Peticiones	T. medio
Validación de cuenta	895	1136388	1270
Consulta de deudas	446	985686	2210
Alta de beneficiario	335	302139	901
Alta de subvención	2130	195631	92
Validar subvención	1676	144548	86

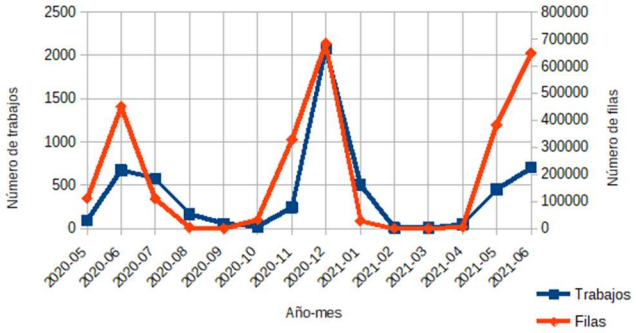


Fig. 8. Trabajos realizados y filas involucradas

GirolA soporta hasta el momento la integración con 19 servicios de todo tipo: síncronos, asíncronos, SOAP, REST, etc. El top 5 de los servicios web más empleados se detalla en la Tabla IV. En esta tabla se muestra el número de trabajos para cada servicio web, el número de filas que afectadas en cada uno de ellos y el tamaño medio de los trabajos. Se aprecia cómo los dos últimos tienen un tamaño medio de lote mucho menor. Ello es debido a la naturaleza de estos servicios. Los servicios de alta y validación de subvenciones son de tipo LCTCS y LCTUS respectivamente (ver sección IV). Se trata por lo tanto de servicios de llamada colectiva que en estos casos estaban limitados a unos 100 integrantes por llamada aproximadamente.

El número de usuarios por perfil existentes y activos en la aplicación se detallan en la Tabla V. Como se puede apreciar a la fecha de la redacción de este artículo el número de usuarios avanzados se ha reducido notablemente. Sólo quedan activos 3. Esto es debido a que a lo largo del ejercicio 2020 mientras los sistemas se integraban con nuestra pasarela muchos de los trabajos se realizaron manualmente y por lo tanto se requería personal para gestionarlos. A lo largo de 2021 las integraciones de varias aplicaciones con GirolA han ido finalizando y sólo unos pocos trabajos necesitan de intervención manual (relacionados con sistemas heredados).

Hasta la fecha se han integrado totalmente con este gateway 3 aplicaciones de la organización:

1. Incentiva: Sistema de gestión de subvenciones y ayudas (Groovy/Grails 1.0.5) Se ha integrado con la pasarela empleando la API de la base de datos.
2. SGTL: Sistema de gestión de residencias de tiempo libre (Oracle 10g + Java). Integrado con GirolA mediante la API de la base de datos.
3. ERTE210: Sistema de gestión de las ayuda ERTE (Oracle APEX). Integrado via servicios web. Integrado con GirolA empleando la API SOAP. Este sistema ha necesitado 20 horas de desarrollo para la integración con nuestro *gateway*.

Tabla V
USUARIOS REGISTRADOS / USUARIOS ACTIVOS

Perfil	Usu. registrados	Usu. activos
Administrador	4	4
U. Avanzado	22	3
Usuario	1	1
Consulta	2	2

Desde un punto de vista cuantitativo la aplicación de este sistema en nuestra organización ha sido un éxito. En apenas un año se han realizado invocaciones a distintos servicios web para 2.784.732 elementos. Se han integrado 19 servicios web distintos que permiten principalmente realizar consultas sobre determinados asuntos, pero también permiten solicitar a los sistemas remotos que elaboren algún trabajo o realicen algunas acciones. Hasta el momento se han integrado 3 aplicaciones con esta pasarela, pero la idea a futuro es que se convierta en un sistema de uso obligatorio por las aplicaciones de la organización.

Desde un punto de vista cualitativo la integración de las aplicaciones con sistemas externos se ha simplificado en enorme medida. Una aplicación antes tenía que integrarse con N sistemas destino. Ahora una aplicación tiene que integrarse con un único sistema: GirolA. Los mensajes que las aplicaciones intercambian con la pasarela son homogéneos. Todos los servicios web de los sistemas destino son tratados de una misma manera cuando se emplea nuestro *gateway*. Hemos podido comprobar que todo ello redunda en una reducción importante del esfuerzo necesario para desarrollar la integración de una aplicación con sistemas externos. Por último, damos soporte a la Integración de aplicaciones heredadas a través del uso de ficheros generados a través de procesos externos a las mismas.

IX. Conclusiones y trabajos futuros

En este artículo hemos presentado GirolA, una pasarela de servicios web cuyos objetivos principales son desacoplar a las aplicaciones de las bibliotecas de integración, simplificar la integración de las mismas con los proveedores de servicios web y resolver el problema de la integración de las aplicaciones heredadas. La experiencia de uso de la pasarela refleja que se han conseguido los objetivos, simplificando significativamente la integración de servicios en una gran organización.

Nuestros planes de futuro con respecto a GirolA pasan por seguir extendiendo su uso en las distintas aplicaciones de la organización. Hasta el momento se ha realizado un esfuerzo en esta línea pero focalizado sólo en algunos sistemas. También deseamos continuar ampliando el número de servicios web soportados por la pasarela y dotarla de una API REST.

Referencias

[1] Singh, Munindar P., and Michael N. Huhns. *Service-oriented computing*. Chichester: Wiley, 2005.

[2] Rashed A. Bahlool, and Ahmed M. Zeki. "Comparative Study between Web Services Technologies: REST and WSDL." *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, p. 1-4, IEEE, 2019.

[3] Tanabe, Yudai, Tomoyuki Aotani, and Hidehiko Masuhara. "A context-oriented programming approach to dependency hell." *Proceedings of the 10th International Workshop on Context-Oriented Programming: Advanced Modularity for Run-time Composition*, pp. 8-14, 2018.

[4] Huy, Hoang Pham, Takahiro Kawamura, and Tetsuo Hasegawa. "Web service gateway-A step forward to e-business." Proceedings. IEEE International Conference on Web Services, 2004, p. 648-655 IEEE, 2004.

[5] "WebSphere Application Server for z/OS web services gateway", https://www.ibm.com/docs/es/was-zos/9.0.5?topic=gateway-web-services-frequently-asked-questions#cwsg_faq__i2

[6] Membrane Service Proxy, https://www.membrane-soa.org

[7] Jan Königsberger and Bernhard Mitschang. "R2SMA-A Middleware Architecture to Access Legacy Enterprise Web Services using Lightweight REST APIs." ICEIS (2), p. 704-711, 2018.

[8] Bixin, Fan Yu Yang Fan Liu, and Zhou Bin. "Research and Implementation of a SOAP-CORBA Gateway System [J]." Computer Engineering and Applications, vol. 29, 2004.

[9] Gerald Brose. "A gateway to web services security–Securing SOAP with proxies." International Conference on Web Services, p. 101-108, Springer, Berlin, Heidelberg, 2003.

[10] Nils Gruschka, and Norbert Luttenberger. "Protecting web services from dos attacks by soap message validation." IFIP International Information Security Conference, p. 171-182, Springer, Boston, MA, 2006.

[11] Ernesto Damiani, et al. "Fine grained access control for SOAP e-services." Proceedings of the 10th international conference on World Wide Web. p. 504-513, 2001.

[12] Katayama Toshiaki, Mitsuteru Nakao, and Toshihisa Takagi. "TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services." Nucleic acids research vol. 38, no suppl_2, p. W706-W711, 2010

[13] SCSP, Sustitución de Certificados en Soporte Papel, https://administracionelectronica.gob.es/ctt/scsp

[14] https://www.agenciatributaria.es/static_files/Sede/Procedimiento_ayuda/ZA05/indice_serv_web_sum_aapp.pdf

[15] jq a lightweight and flexible command-line JSON processor, https://stedolan.github.io/jq

[16] Antonio Delgado, et al. "Reusing UI elements with model-based user interface development." International Journal of Human-Computer Studies, vol. 86, p. 48-62, 2016

[17] Shashank Sharma and Thomas Keir. "Scheduling Tasks." Beginning Fedora: From Novice to Professional, p. 427-431, 2007


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Hacia la anotación y realización de tareas de aprendizaje ubicuo en el contexto de historia del arte

Pablo García Zarza, Adolfo Ruiz Calleja, Miguel L. Bote Lorenzo
Guillermo Vega Gorgojo, Eduardo Gómez Sánchez, Juan I. Asensio Pérez
Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática
ETSI de Telecomunicación, Universidad de Valladolid
Campus Miguel Delibes, Paseo de Belén 15, 47011 Valladolid.
{pablogz, adolfo}@gsic.uva.es, {migbot, guiveg, edugom, juaase}@tel.uva.es

LocalizARTE es una aplicación distribuida para publicar y realizar actividades educativas relacionadas con historia del arte donde la información que utiliza de partida ha sido generada a partir de datos abiertos ofrecidos por distintas organizaciones y las anotaciones que se realicen se proporcionarán también como datos abiertos. Su objetivo es apoyar al aprendizaje en diferentes espacios físicos y virtuales. En este artículo se ilustrará la ontología utilizada para las anotaciones y el funcionamiento, la arquitectura y cómo se está implementando la aplicación. Se describirá a través de un escenario de ejemplo donde un profesor de Historia del Arte publica nuevas tareas educativas en LocalizARTE con el objetivo de que sus estudiantes visiten y analicen los monumentos y edificios de su entorno.

***Palabras Clave*—Datos Abiertos Enlazados, anotación semántica, aplicaciones distribuidas, aprendizaje ubicuo**

I. INTRODUCCIÓN

El aprendizaje ubicuo se puede definir como «utilizar las tecnologías móviles para facilitar el aprendizaje», si bien la definición «aprender en cualquier lugar y en cualquier momento» [1] es la que se usa con mayor frecuencia. Ambas definiciones plantean que las situaciones de aprendizaje ubicuo ocurren en distintos espacios físicos (p. ej. una clase o un museo) o virtuales (p. ej. un entorno de aprendizaje virtual o una aplicación móvil) siendo los dispositivos móviles los que dan soporte a ese aprendizaje [1], [2]. La comunidad investigadora y docente ha mostrado un claro interés por el aprendizaje ubicuo en los últimos años debido a que otorga autonomía al estudiante y favorece el aprendizaje situado [2], [3].

El aprendizaje ubicuo es especialmente importante para el campo de Historia del Arte porque los estudiantes adquieren una mayor comprensión de un monumento al visitarlo y analizarlo presencialmente, en comparación con estudiarlo a través de libros o recursos web [4]. Debido a ello, en trabajos anteriores hemos propuesto Casual Learn [5], [6], una aplicación móvil que da soporte al aprendizaje ubicuo de historia del arte y que actualmente está disponible en Google Play[1]. Casual Learn recomienda la realización de tareas de aprendizaje relacionadas con monumentos cuando se pasa cerca de ellos. Como característica distintiva, Casual Learn se basa en tecnologías semánticas y de la Web de Datos Abiertos [7]. Esto ha permitido generar semiautomáticamente 10 000 tareas geolocalizadas en monumentos de Castilla y León que se estructuran según la ontología SLEek [8] y se ofrecen como datos abiertos en un punto SPARQL[2]. Todas estas tareas se definieron teniendo en cuenta el currículum educativo de esta comunidad autónoma y fueron validadas por profesores de Educación Secundaria Obligatoria [9].

A pesar de que dichas tareas representan una excelente colección para Casual Learn, adolecen del problema de que los datos utilizados para generarlas no describen particularidades históricas, simbólicas u ornamentales de edificios civiles [9]. Esta carencia puede ser subsanada permitiendo la publicación de nuevos datos a expertos de su comunidad de usuarios. Además, otra limitación de estas tareas es que están diseñadas para ser realizadas exclusivamente en el espacio físico. Por ello, si se desea facilitar la creación de puentes entre distintos espacios educativos (p. ej., entre el espacio físico y un entorno virtual de aprendizaje como Moodle) surge la necesidad de permitir la creación de tareas específicas para otros espacios que no tengan en cuenta la ubicación del usuario.

Para superar estos dos problemas proponemos LocalizARTE, una aplicación para publicar y realizar tareas de historia del arte tanto en el espacio físico como en el virtual. Constará de una aplicación web con dos interfaces

[1] https://casuallearnapp.gsic.uva.es
[2] https://casuallearn.gsic.uva.es/sparql

basadas en mapas. La primera de ellas permitirá a los estudiantes resolver tareas en el espacio físico de forma similar a como lo permite Casual Learn, además de posibilitarles completar otras tareas en el espacio virtual a través del mapa (es decir, sin necesidad de estar físicamente en las proximidades de un monumento). Con la segunda interfaz los profesores podrán crear nuevas tareas y evaluar las realizadas por sus alumnos. Esto supone un reto, ya que LocalizARTE deberá permitir a usuarios no técnicos (docentes) la publicación en la Web de Datos de tareas geolocalizadas en diferentes puntos físicos y virtuales. Para poder ilustrar su funcionamiento, como datos de partida específicos para esta aplicación se han generado 3600 tareas de aprendizaje para ser mostradas en el mapa virtual (generadas de forma similar a [10]).

Bien es cierto que existen otras aplicaciones que dan soporte a la publicación, desarrollo y evaluación de tareas geolocalizadas [11], [12], [13]. En estas aplicaciones la generación de tareas se basa en una comunidad de creadores de contenido partiendo de un registro vacío y no son compartidas entre ellas. En el caso de LocalizARTE se utilizan tecnologías de la Web de Datos para publicar de manera abierta las tareas generadas siguiendo una aproximación social y semántica que parte de datos ya publicados en la Web de Datos. Este tipo de tecnologías –que ya han sido empleadas en el dominio educativo [14]– suponen una mayor dificultad, ya que se debe ofrecer al docente (generalmente sin experiencia en tecnologías semánticas) una interfaz de publicación que esconda la complejidad de la estructura de datos subyacente.

II. LA APLICACIÓN LOCALIZARTE

A. Ejemplo de escenario de uso

Rodrigo es un profesor de Historia del Arte que ejerce en un instituto de Educación Secundaria de Valladolid. Tras explicar el arte gótico, considera que sería bueno que sus estudiantes visitasen algunos de los edificios góticos vallisoletanos. De esta forma comprenderán y respetarán el patrimonio local a la vez que reforzarán los contenidos vistos en clase. Para ello sugerirá a sus estudiantes que utilicen LocalizARTE.

Antes de hablar con ellos Rodrigo consulta las **tareas** disponibles en LocalizARTE. Comprueba que todos los edificios góticos relevantes de Valladolid han sido identificados como puntos de interés (**POI**, *point of interest*), están descritos y contienen algunas tareas que les serían interesantes. Sin embargo, decide añadir otras tareas para subrayar algunos de los contenidos trabajados en clase. Así, añade una tarea que invita a la reflexión sobre la iconografía del Colegio de San Gregorio y otra que solicita comparar las ventanas góticas de la iglesia de San Pablo con las de la torre románica de la iglesia de San Martín.

Rodrigo también es consciente de que Valladolid tiene edificios muy importantes del gótico isabelino, pero carece de edificios de un gótico más primitivo. Por ello, consideraría buena idea que sus estudiantes realizasen tareas geolocalizadas en otras ciudades de la región, como León o Burgos. Observa que en LocalizARTE hay definidos

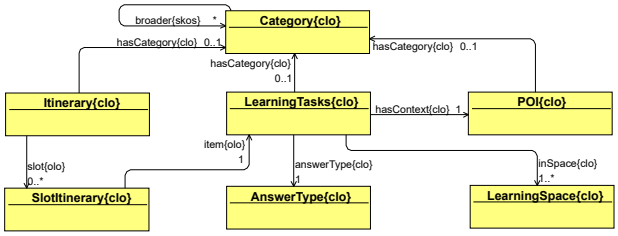


Fig. 1. Visión general de la ontología de LocalizARTE.

bastantes POI con tareas interesantes en esas ciudades que se pueden realizar desde el mapa virtual. Algunas parecen propuestas por profesores de esas mismas ciudades. Con todo, Rodrigo propone a sus alumnos que utilicen LocalizARTE para realizar tareas en cuatro edificios góticos de Valladolid, siendo obligatorio visitar el Colegio de San Gregorio y la iglesia de San Pablo (**itinerario**). También deben realizar tareas en el mapa virtual sobre los edificios góticos de otra ciudad de Castilla y León. A partir de las tareas realizadas (algunas presencialmente, en el espacio físico, y otras virtualmente), los estudiantes entregarán una comparativa de la arquitectura gótica (**portafolio**).

B. Anotaciones soportadas

Los datos de partida que van a ser empleados en este proyecto son los disponibles en Casual Learn SPARQL[2]. Esta información se ha generado a partir de datos del portal de Datos Abiertos de la Junta de Castilla y León, esDBpedia y Wikidata [5]. Los datos de este repositorio se dividieron en tres clases: POI, tareas de aprendizaje y temas. LocalizARTE utilizará una versión extendida de esta ontología cuya estructura general se puede apreciar en la Fig. 1. En la ontología de LocalizARTE se agregan los espacios en los que se pueden realizar las tareas (físico si se tiene que realizar cerca de donde esté ubicado el POI, mapa virtual si se puede realizar desde cualquier lugar a través de un mapa) e itinerarios (serie de tareas de distintos POI). En el Listado 1 se puede ver cómo LocalizARTE llevaría a cabo la inserción en el repositorio, utilizando SPARQL Update [15], de una de las tareas propuestas en el apartado A. LocalizARTE permite a los estudiantes realizar distintos tipos de tareas, lo que influye en el tipo de respuesta esperado (en el ejemplo del Listado 1 la respuesta esperada es una fotografía acompañada de un texto). La gestión de los datos abiertos utilizados en LocalizARTE recaerá sobre su autor, teniendo límites. Por ejemplo, un docente podrá eliminar uno de sus POI siempre y cuando no tenga ninguna tarea asociada.

Listado 1. Inserción de una tarea en el grafo de LocalizARTE

```
with <http://localizarte.gsic.uva.es> insert {
  cld:ColegiodeSanGregorio_describeiconografía
    a clo:LearningTask ;
    clo:answerType clo:PhotoTextAnswerType ;
    clo:hasCategory dbc:Isabelline_architecture ;
    clo:hasContext cld:ColegiodeSanGregorio_4723799 ;
    clo:inSpace clo:PhysicalSpace ;
    rdfs:comment "Realiza una fotografía [...]"@es ;
    rdfs:label "Describe la iconografía"@es .
}
```

Fig. 2. Recortes de pantalla de LocalizARTE Web. (a) POI disponibles en Valladolid junto con sugerencias para crear un nuevo POI por parte de un docente. Los POI se pueden agrupar en marcadores circulares dependiendo del nivel de zum y de la distancia entre ellos. El número del marcador indica los puntos agrupados. (b) Anotación de una tarea para el POI Colegio de San Gregorio como se indica en el caso de uso de ejemplo.

Tabla I
PREFIJOS Y NAMESPACES UTILIZADOS EN ESTE DOCUMENTO.

Prefijo	*Namespace*
clo:	https://casuallearn.gsic.uva.es/ontology/
cld:	https://casuallearn.gsic.uva.es/data/
dbc:	http://dbpedia.org/resource/Category:
olo:	http://smiy.sourceforge.net/olo/spec/orderedlistontology.html#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
skos:	http://www.w3.org/2004/02/skos/core#

C. *Arquitectura software*

Los usuarios no técnicos pueden tener problemas para ejecutar operaciones como la del Listado 1 por lo que es necesario proporcionarles elementos en la aplicación (Fig. 3) que les faciliten realizar estas acciones. De este modo, siguiendo con el escenario de ejemplo, el profesor Rodrigo accedería al sistema a través de **LocalizARTE Web**, aplicación adaptable que se ejecutará sobre el navegador web. Con ella visualizará los POI y tareas existentes en el sistema a través de un mapa (información recuperada de **LocalizARTE Data** a través de **LocalizARTE Server**). Si decide agregar un POI se le sugerirán una serie de monumentos cercanos al lugar donde quiera añadirlo (Fig. 2a). La información de estas sugerencias se obtiene de datos abiertos de los repositorios de la versión internacional y española de DBpedia. Si decide crear un POI basado en una sugerencia se enlazarán

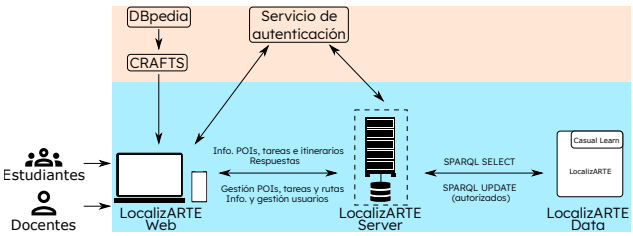


Fig. 3. Arquitectura de LocalizARTE. Sobre un fondo azul se colocan las partes en desarrollo. En naranja los servicios y repositorios externos.

los nuevos datos de LocalizARTE con los originales. Como en este caso Rodrigo solo quiere agregar una serie de tareas, se le mostrará un formulario como el de la Fig. 2b al pulsar sobre una opción del POI del Colegio de San Gregorio. La solicitud de creación de la tarea llegará a **LocalizARTE Server** a través de una API REST [16] donde se realizará una comprobación de los datos introducidos y de la identidad del usuario. Si todo es correcto, se formará una petición de inserción como la del Listado 1 que se enviará a **LocalizARTE Data** para ser almacenada dentro del grafo de LocalizARTE. Tanto este grafo como el de Casual Learn utilizan la estructura de datos de la ontología SLEek [8].

Los estudiantes de Rodrigo pueden comenzar a realizar las tareas en cuanto las haya terminado de agregar. Como ejemplo, una alumna decide comprobar si su profesor ha agregado nuevas tareas. Para ello se identifica en **LocalizARTE Web** y observa que se ha creado la tarea de la Fig. 2b. Al igual que sucedía en el caso del profesor, **LocalizARTE Web** obtiene estos datos de **LocalizARTE Data** a través de **LocalizARTE Server** con el objetivo de mantener en un único lugar las tecnologías de la Web Semántica. Antes de realizar esta tarea, que precisa encontrarse cerca de la ubicación del POI (espacio físico), la alumna realiza alguna de las otras tareas que puede responder desde cualquier lugar con su ordenador (espacio mapa virtual). Con estas tareas resueltas, decide dar un paseo por el centro de la ciudad para llegar hasta la ubicación del POI y cuando está cerca completa con su teléfono móvil la tarea que ha creado Rodrigo.

Las respuestas de la alumna se almacenan en una base de datos privada de **LocalizARTE Server**. Solo ella y su profesor podrán ver las respuestas que ha proporcionado a través de un portafolio. Para ello, la alumna deberá haberse registrado dentro de uno de los cursos de su profesor en LocalizARTE. Rodrigo, además de poder visualizar las respuestas, podrá proporcionar realimentación a su alumna desde la interfaz gráfica de la aplicación.

D. Implementación

Puesto que se desea que el cliente de LocalizARTE pueda utilizarlo todo tipo de usuarios se optó por un desarrollo con tecnologías nativas de la Web para que pueda ser ejecutado sobre navegadores web. Para el uso básico de LocalizARTE (búsqueda de información) la interfaz deberá mostrar los POI como una serie de marcadores interactivos sobre un mapa. Al pulsar sobre uno de ellos se mostrará su información junto con las tareas que están agrupadas en ese POI. Estas dos acciones se consiguen solventar a través de las bibliotecas `Leaflet` (v1.7.1, utilizando el *plugin* `markercluster`[3]) y `Bootstrap` (v5.0.1, que permite generar una interfaz adaptable al dispositivo). Los mapas se basan en datos de OpenStreetMap, habiendo creado un diseño personalizado con mapbox[4] para que solo se muestren calles y edificios (evitando distracciones provocadas por otros iconos). El resto de la lógica del cliente se está desarrollando con JavaScript.

Como el cliente va a estar gestionando y solicitando objetos nativos de JavaScript, LocalizARTE Server también se está implementando usando tecnologías de la Web para manipular los mismos tipos de datos. Como entorno de ejecución de JavaScript se está utilizando Node.js que permite la creación de una API REST a través de `Express` y una instancia de MongoDB para almacenar la información que generen los usuarios.

LocalizARTE Data utiliza Virtuoso Open Source Edition[5] (v7.2.5) como *middleware* para poder exponer los datos. Desde LocalizARTE Server se crearán, modificarán o eliminarán datos de LocalizARTE utilizando SPARQL Update. Para las comunicaciones con el resto de repositorios se utiliza CRAFTS[6], que permite simplificar el acceso a la información almacenada en repositorios de triplas.

III. CONCLUSIONES

LocalizARTE es una aplicación que permite anotar y realizar tareas para aprender historia del arte publicadas como datos abiertos en la Web. Esta aplicación posibilita que cualquier docente pueda generar nuevos contenidos sin que sea un experto en tecnologías de la Web Semántica lo que supone una ventaja frente a otras aplicaciones como Casual Learn. Además, es compatible con actividades educativas de varios espacios lo que facilita la construcción de puentes entre los contenidos vistos en clase y el patrimonio cultural del entorno de los estudiantes. Su arquitectura distribuida facilita el acceso ubicuo y controlado a los datos que existan y se generen en la aplicación. Una vez finalizado su desarrollo publicaremos LocalizARTE de manera abierta habiendo realizado previamente un análisis de la capacidad de carga del sistema y su grado de usabilidad. Además, se llevarán a cabo varios proyectos pilotos en centros de Educación Secundaria de la comunidad autónoma de Castilla y León.

[3]https://github.com/Leaflet/Leaflet.markercluster
[4]https://mapbox.com
[5]https://github.com/openlink/virtuoso-opensource
[6]https://crafts.gsic.uva.es

AGRADECIMIENTOS

La investigación aquí reportada ha sido financiada por el FEDER y la Consejería de Educación de la Junta de Castilla y León bajo el proyecto VA257P18, y el Fondo Europeo de Desarrollo Regional y la Agencia Nacional de Investigación del Ministerio de Ciencia e Innovación bajo el proyecto TIN2017-85179-C3-2-R.

REFERENCIAS

[1] G. J. Hwang y C. C. Tsai, "Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010," *British Journal of Educational Technology*, vol. 42, no. 4, pp. E65–E70, 2011.

[2] G. Pishtari, M. Rodríguez-Triana, E. Sarmiento-Márquez, M. Pérez-Sanagustín, A. Ruiz-Calleja, P. Santos, L. Prieto, S. Serrano-Iglesias, y T. Väljataga, "Learning design and learning analytics in mobile and ubiquitous learning: A systematic review," *British Journal of Educational Technology*, vol. 51, no. 4, pp. 1078–1100, 2020.

[3] L. Cárdenas-Robledo y A. Peña-Ayala, "Ubiquitous learning: A systematic review," *Telematics and Informatics*, vol. 35, no. 5, pp. 1097–1132, 2018.

[4] J. Greene, B. Kisida, y D. Bowen, "The educational value of field trips: taking students to an art museum improves critical thinking skills, and more," *Education Next*, vol. 14, no. 1, pp. 78–86, 2014.

[5] A. Ruiz-Calleja, M. Bote-Lorenzo, G. Vega-Gorgojo, S. Serrano-Iglesias, P. García-Zarza, J. Asensio-Pérez, y E. Gómez-Sánchez, "Casuallearn: A smart application to learn history of art," in *Proceedings of the European Conference on Technology Enhanced Learning (ECTEL)*. Heidelberg, Germany: Springer, 2020, pp. 472–476.

[6] A. Ruiz-Calleja, P. García-Zarza, G. Vega-Gorgojo, M. Bote-Lorenzo, E. Gómez-Sánchez, J. Asensio-Pérez, S. Serrano-Iglesias, y A. Martínez-Monés, "Casual Learn: A semantic mobile application for learning local Cultural Heritage," *Semantic Web Journal*, en revisión.

[7] T. Heath y C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.

[8] A. Ruiz-Calleja, M. Bote-Lorenzo, G. Vega-Gorgojo, A. Martínez-Monés, J. Asensio-Pérez, E. Gómez-Sánchez, S. Serrano-Iglesias, y Y. Dimitriadis, "SLEek: An Ontology For Smart Learning in the Web of Data," en *Proceedings of the 21st IEEE International Conference on Advanced Learning Technologies*. Texas, USA: ACM, 2021.

[9] A. Ruiz-Calleja, G. Vega-Gorgojo, M. Bote-Lorenzo, J. Asensio-Pérez, Y. Dimitriadis, y E. Gómez-Sánchez, "Supporting contextualized learning with linked open data," *Journal of Web Semantics*, vol. 70, p. 100657, 2021.

[10] A. Ruiz-Calleja, M. Bote-Lorenzo, J. Asensio-Pérez, G. Vega-Gorgojo, Y. Dimitriadis, A. Martínez-Monés, E. Gómez-Sánchez, y S. Serrano-Iglesias, "Automatic creation of Moodle activities out of the Web of Data to link formal and informal learning contexts," en *Proceedings of the 8th International Conference on Technological Ecosystems for Enhancing Multiculturality*. Salamanca, Spain: ACM, 2020, pp. 238–244.

[11] T. Väljataga y K. Mettis, "Turning zoos into smart learning ecosystems," *Interaction Design and Architecture(s) Journal*, vol. 39, no. 39, pp. 114–133, 2016.

[12] S. Yu, X. Yang, G. Cheng, y M. Wang, "From learning object to learning cell: A resource organization model for ubiquitous learning," *Educational Technology & Society*, vol. 2, no. 18, pp. 206–224, 2014.

[13] E. FitzGerald, "Creating user-generated content for location-based learning: an authoring framework," *Journal of Computer Assisted Learning*, vol. 3, no. 28, pp. 195–207, 2012.

[14] G. Vega-Gorgojo, J. Asensio-Pérez, E. Gómez-Sánchez, M. Bote-Lorenzo, J. Muñoz-Cristóbal, y A. Ruiz-Calleja, "A review of linked data proposals in the learning domain," *Journal of Universal Computer Science*, vol. 21, pp. 326–364, 2015.

[15] P. Gearon, A. Passant, y A. Polleres, *SPARQL 1.1 Update*. Disponible en: https://www.w3.org/TR/sparql11-update/, última visita en junio 2021

[16] R. T. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, 2000.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Enhancing rescue operations with virtualized mobile services in scarce resource devices

Asier Atutxa, Jasone Astorga, Maider Huarte, Eduardo Jacob, Juanjo Unzilla
Department of Communications Engineering,
University of the Basque Country (UPV/EHU), Faculty of Engineering,
Alda. Urquijo S/N, 48013 Bilbao, Spain.
asier.atutxa@ehu.eus, jasone.astorga@ehu.eus, maider.huarte@ehu.eus, Eduardo.Jacob@ehu.eus, juanjo.unzilla@ehu.eus

***Abstract*—This paper presents [1] an architecture to support reconfigurable multimedia services for a practical emergency environment of rescue operations. This solution is validated by implementing a digital mobile radio (DMR) standard radio hotspot and a video streaming server in an unmanned aerial vehicle (UAV) that conveys a device with scarce resources in order to minimize power consumption. To achieve a fast and flexible deployment of the envisioned services, a virtualization-based approach using a Kubernetes orchestrator is used, which manages the life-cycle of services. Results show that the proposed architecture can be deployed in less than 4 minutes and can recover from network disconnections in less than 10 seconds.**

***Keywords*—Application virtualization, Emergency services, Multimedia communication, Radio communication.**

I. PROPOSED ARCHITECTURE

The first design objective must be to allow providing rescue teams with radio communications and live video streaming of the affected area as soon as possible. To achieve this goal, we have chosen to use an unmanned aerial vehicle (UAV) to carry the infrastructure, as it is able to reach the targeted area sooner than any other vehicle. However, the use of UAVs implies some restrictions on the equipment that can be carried, because it must be light and must consume little power. Therefore, we chose to use a Raspberry Pi for the deployment of the digital mobile radio (DMR) [2] and video streaming services using UDP.

Another important requirement is that the solution must be reconfigurable in a fast and flexible manner. We have opted for a virtualization-based approach, based on Docker containers. These containers provide environments tailored to a specific purpose, using suitable packages and libraries. Furthermore, particular configurations can be specified in the deployment phase, such as radio frequency parameters. Fig. 1 provides a global view of the proposed architecture design.

II. RESULTS AND DISCUSSION

Table I presents a summary of the performed time-measurement tests. Results show that the most time consuming process is the Kubernetes cluster start up time, which needs 1 minute and 51,6 seconds on average. However, deploying a service needs 19,3 seconds for the video streaming and 17 seconds for the DMR communication, which is a competitive result.

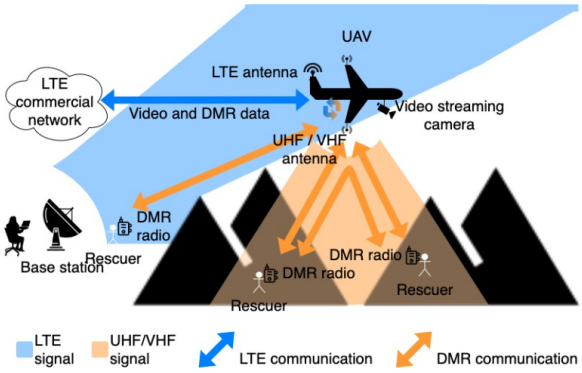


Fig. 1. Proposed architecture.

Table I
SUMMARY OF THE RESULTS OF THE PERFORMED TESTS.

Measured Time	Mean Time	95% Conf Interval
K8s cluster start up	1 min 51.6 s	1 min 50.4 s - 1 min 52.8 s
Cluster node start up	1 min 42.0 s	1 min 38.8 s - 1 min 45.1 s
Video service start up	19.3 s	18.5 s - 20.0 s
DMR service start up	17.0 s	16.5 s - 17.4 s
Node reconnection	8.0 s	7.6 s - 8.3 s

ACKNOWLEDGEMENTS

This work was supported in part by the TEC2016-76795-C6-5-R, and in part by the CogNoms4.0 KK-2018/00049 research project.

REFERENCES

[1] A. Atutxa, J. Astorga, M. Huarte, E. Jacob and J. Unzilla, "Enhancing Rescue Operations With Virtualized Mobile Multimedia Services in Scarce Resource Devices," in IEEE Access, vol. 8, pp. 216029-216042, 2020, doi: 10.1109/ACCESS.2020.3041394.

[2] ETSI, ETSI TS 102 361-3 V1.1.7 (2007-12), Electromagnetic compatibility and Radio spectrum Matters (ERM): Digital Mobile Radio (DMR) Systems Part 3: DMR data protocol, December 2007.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Análisis del despliegue de servicios de misión crítica en el extremo de la red

Aitor Sanchoyerto[1], Begoña Blanco[1], Endika Aldecoa[1], Fidel Liberal[1]

[1] University of the Basque Country, Bilbao (Spain)

{aitor.sanchoyerto, bego.blanco, endika.aldecoa,fidel.liberal}@ehu.eus[1]

Los usuarios de las agencias PPDR (Public Protection Disaster Recovery) han identificado la conveniencia de utilizar recursos multimedia para complementar a la voz en comunicaciones de misión crítica. Estas funcionalidades disponibles en las redes celulares de banda ancha de propósito general, no lo estaban en las redes de comunicaciones que emplean tradicionalmente las agencias PPDR (TETRA, TETRAPOL, P25), debido a sus limitadas capacidades de transmisión de datos. El trabajo llevado a cabo en los últimos años dentro del grupo SA6 del 3GPP (3rd Generation Partnership Project), ha permitido incluir nuevas funcionalidades y capacidades al estándar LTE, garantizando la idoneidad de uso por parte de las agencias PPDR. La estandarización de los servicios de misión crítica: Mission Critical Push to Talk (MCPTT), Mission Critical Data (MCDATA) y Mission Critical Video (MCVIDEO) habilitan el uso para misión crítica de las redes de banda ancha, como Mission Critical Broadband (MBB). Las redes MBB ofrecen a las agencias PPDR el tradicional servicio de voz y dan cumplimiento a los nuevos requerimientos (video, imágenes o transferencia de datos). Durante una emergencia, se produce un aumento exponencial del número de llamadas de emergencia en la MBB. El mayor número de éstas se concentran en el área geográfica donde se ha identificado el incidente. La red debe estar dimensionada para gestionar este aumento de capacidad y garantizar los indicadores de servicio KPIs para todas las llamadas. A priori, el despliegue del servicio de misión crítica MCPTT en el extremo de la red donde se haya producido la emergencia parece ser la acción a seguir por el gestor de la red de cara a garantizar los niveles de servicio. El simulador desarrollado permite analizar cómo afecta este despliegue a los indicadores de servicio KPIs del conjunto de llamadas de emergencia que se están estableciendo en la red MBB. El simulador de despliegue de servicios MCPTT desarrollado permite comparar diferentes escenarios desde el punto de vista de los indicadores de servicio KPI, definidos por el 3GPP para el conjunto de llamadas establecidas en la red de misión crítica.

***Palabras Clave*- PPDR, MCPTT, NFV, EDGE**

I. INTRODUCCIÓN

El National Public Safety Telecommunications Council (NPSTC) en su informe "Push to Talk over Long Term Evolution Requirements" [1] enumera un conjunto de requerimientos (ver Tabla I) que debe cumplir el servicio PTT (Push to Talk) al ser desplegado sobre redes de banda ancha.

Esta relación de requerimientos funcionales [2] está basada en el servicio PTT que ofrecían las redes PMR/LMR y en la experiencia de usuarios que diariamente trabajan con esta tecnología.

Tabla I
Requerimientos funcionales según NPSTC [8]

Naturaleza	Descripción
Llamada	Llamada a grupo, llamada privada y llamada a grupo con aceptación de llamada
Servicio	Rendimiento de las llamadas PTT, PTT Late Call Entry, gestión de grupos dinámicos y monitorización de las llamadas.
Prioridad	Priorización de las llamadas y llamadas de emergencia.
Seguridad e Identificación	*Identificación de la llamada, gestión de perfiles, seguridad y localización.*

El objetivo del NPSTC [3] ya en 2013 era que estos requerimientos fueran recogidos por el 3GPP (Third Generation Partnership Project) en las futuras versiones del estándar de LTE. Y así ocurrió cuando a finales de 2016 quedó reflejado en la Release 13. El 3GPP, para poder cuantificar este rendimiento, ha definido un indicador para cada una de las fases de una llamada de misión crítica PTT: establecimiento, mantenimiento e inclusión en una llamada existente. Su definición, así como los retardos que se deben cumplir y el porcentaje de llamadas que lo deben cumplir se encuentran indicados en 3GPP TS 22.179 [4]

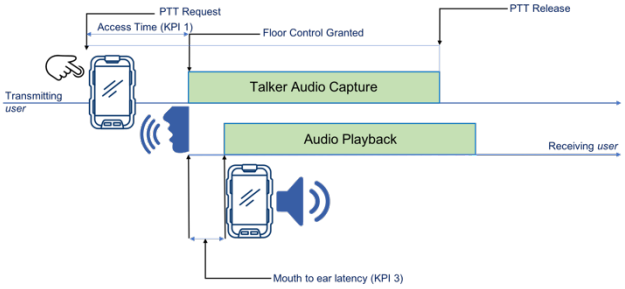


Fig. 1: MCPTT Access Time and Mouth-to-Ear Latency (Based on [4])

Los indicadores de rendimiento definidos por el 3GPP [4] que se van a analizar en este trabajo son los siguientes:

- **KPI_1 – Access Time.** Se define como el tiempo entre el momento en que un usuario de MCPTT solicita hablar (presionando el botón MCPTT en el terminal móvil de usuario de MCPTT (UE)) y el momento en que este usuario recibe la autorización para hablar. Este tiempo no incluye las confirmaciones de los usuarios receptores o el tiempo para afiliarse al grupo.
- **KPI_2: Access Time End to End.** Se define como el tiempo entre el momento en que un usuario MCPTT solicita hablar y el momento en que este usuario recibe una señal para comenzar a hablar, incluido el establecimiento de la llamada MCPTT (si corresponde) y el reconocimiento del primer receptor/usuario antes de que se pueda transmitir la voz.
- **KPI_3: Mouth to Ear Latency**. Es el tiempo que transcurre desde que el emisor habla y la reproducción de la emisión en el altavoz del usuario receptor.

El 3GPP en TS 22.179 V14.3.0 (2016-12) [4] define la latencia máxima medida en milisegundos para cada uno de los indicadores de servicio descritos anteriormente y el porcentaje de llamadas que deben de cumplir estos tiempos máximos. En la tabla siguiente se muestran cada uno de estos valores para los 3 KPIs más relevantes:

Tabla II
Threshold defined for MCPTT KPIs

MCPTT KPI	Threshold	Likelihood	LTE Packet Delay Jacket
KPI_1 – Access Time	< 300 ms	95% of all MCPTT requests	< 60 ms
KPI_2 End to End Access Time	< 1000 ms	N/A	< 60 ms
KPI_3 Mouth to Ear Latency	< 300 ms	95% of all voice bursts	< 75 ms

El cumplimiento de estos requisitos depende de la arquitectura de la red sobre la que se despliegue el servicio y del propio servicio.

II. SIMULADOR DE DESPLIEGUE MCPTT

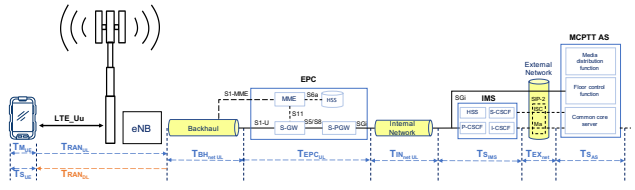


Fig. 2: Identificación de tiempos de proceso y red, en una arquitectura estándar LTE.

Se ha elegido una arquitectura LTE estándar versión 13, como modelo de referencia de red de banda ancha sobre la cual desplegar el servicio de misión crítica MCPTT. Esta red está compuesta de: una red de acceso radio (RAN) compuesta por las estaciones base (eNB_1, eNB_2) y el conjunto de usuarios (X_1, X_2), un núcleo de red (EPC_1, EPC_2) y un Data Center (MAIN) con un IP Multimedia Subsystem (IMS) que es SIP Core sobre el que se despliega el servicio MCPTT en condiciones de funcionamiento estándar de la red de emergencia. En el supuesto de desplegar el servicio en el extremo de la red ($eNB_{1,2}$) se realizará junto con el SIP CORE.

De cara a evaluar de forma sencilla el valor de cada uno de los indicadores de servicios KPI en función del destinatario final de la llamada MCPTT se han analizado las contribuciones de los diferentes componentes de las arquitecturas LTE a cada KPI [5]. A partir de esos cálculos se ha desarrollado un entorno de simulación que permite analizar la evolución de esos indicadores. La forma más gráfica de describir el simulador es a través de su pantalla principal. Esta utilidad ha sido desarrollada en Matlab y constituye una herramienta de análisis gráfico de gran potencia.

Se han definido tres escenarios de simulación que actualmente son fijos pero que en futuras versiones de esta herramienta se está analizando la posibilidad de parametrizar el modelizado de los escenarios a través de ficheros de texto plano que puedan fácilmente ser importados por la herramienta al inicio de la ejecución. Esto permitirá sin duda una mayor flexibilidad a los análisis futuros.

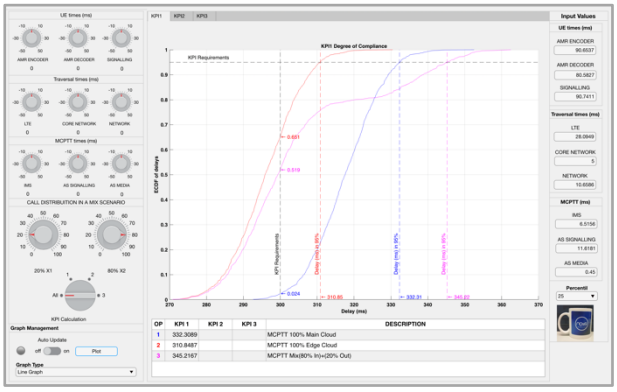


Fig. 3: Pantalla principal del simulador de despliegue del servicio MCPTT

Para favorecer la toma de decisiones se puede realizar la simulación de cada uno de los escenarios, para lo cual se emplean los tiempos de proceso o de red (latencia) medido en cada uno de los componentes que intervienen en la llamada: [5]

- UE times (ms): AMR ENCODER, AMR DECODER, SIGNALLING
- Traversal times (ms): LTE, BACKHAUL, EPC, NETWORK (Internal/External)
- MCPTT times (ms): IMS, AS SIGNALLING, AS MEDIA

Estos tiempos han sido obtenidos a partir del procesamiento en Matlab de las trazas capturadas con la herramienta Wireshark correspondientes a 1000 llamadas MCPTT, con una cadencia de 20 llamadas por segundo. A continuación, se detalla la formulación que se ha implementado en el simulador para el cálculo de cada uno de los indicadores de servicio. [5].

$$KPI_1 = TRAN_{UL} + 2\,(TS_{UE1} + TBH_{net} + TEPC + TIN_{net} + TS_{IMS} + TEX_{net} + TS_{AS}) + TRAN_{DL}$$

$KPI_2 = 2\ (T_{RANUL} + T_{SUE1} + T_{SAS}) + 4(T_{BHnet} + T_{EPC} + T_{INnet} + T_{SIMS} + T_{EXnet}) + T_{SUE2} + 2\ T_{RANUL}$

$KPI_3 = T_{RANUL} + T_{MUE1} + 2\ (T_{BHnet} + T_{EPC} + T_{INnet} + T_{EXnet}) + T_{MAS} + T_{MUE2} + T_{RANDL}$

donde:

Tabla III

Tiempos de Procesado

Tiempo	Descripción
T_{SUE}	Señalización en el cliente
T_{MUE}	Datos media en el cliente UE
T_{SIMS}	Señalización IMS
T_{SAS}	Señalización Servidores MCPTT
T_{MAS}	Datos media Servidores MCPTT

Tabla IV

Tiempos de Retardo

Tiempo	Descripción
T_{RANUL}	Tiempo E-UTRAN uplink
T_{RANDL}	Tiempo E-UTRAN downlink
T_{BHnet}	Tiempo backhaul
T_{INnet}	Tiempo de internal network
T_{EXnet}	Tiempo de external network

A. Objetivo Principal

Analizar el efecto del despliegue del servicio de misión crítica MCPTT en el extremo de la red en los indicadores de servicio KPIs del conjunto de llamadas de la red MBB.

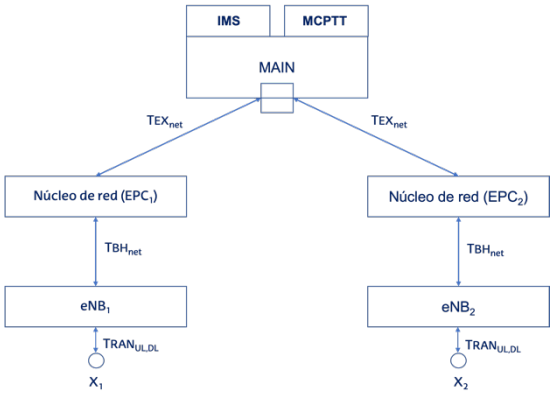


Fig. 4. OPCION 1 - MCPTT desplegado en el Data Center del Operador

El escenario base modelizado en el simulador es:

- Arquitectura estándar LTE en su versión 13.
- Data Center del operador de red, denominado MAIN.
- El core IMS se despliega junto al servicio de misión crítica MCPTT en el Data Center del operador o en el extremo de la red.
- Las estaciones Base (eNB_1 y eNB_2) son los extremos de la red a través de las cuales se conectan los usuarios X_1 y X_2.
- $X_1 + X_2$ representan los usuarios que participan en una llamada MCPTT en el instante de tiempo de simulación
- X_1, X_2 pueden acceder a la red dependiendo de su ubicación geográfica a través de eNB_1 y eNB_2.
- eNB_1 y eNB_2 se conectan al MAIN a través del núcleo de red (EPC_1, EPC_2)
- T_{EXTnet} representa el retardo que introduce la red externa, que conecta el núcleo de red (EPC_1, EPC_2) con el MAIN.
- T_{INTnet} se considera despreciable al desplegar juntos el servicio MCPTT y el core IMS. [6]
- A efectos de la simulación, el retardo que añade el backhaul es T_{BHnet} y el que añade la red de acceso radio (RAN) es T_{RANUL} en el uplink y T_{RANDL} en el downlink
- El servicio MCPTT se despliega en el MAIN o en uno de los extremos de la red (EDGE): eNB_1 o eNB_2. Se ha elegido el nodo eNB_2, para desplegar el servicio MCPTT y el Core IMS.

B. Escenarios de Despliegue

Los tres escenarios de despliegue que se contemplan en el simulador son los siguientes:

- Opción1: El servicio MCPTT está desplegado en el MAIN, Fig. 4. Los usuarios X_1, X_2 acceden al servicio centralizado que se encuentra en el Data Center del operador. Se ha considerado que los usuarios X_1 se conectan a eNB_1 y los X_2 a través de eNB_2, aunque podría conectarse X_1 a eNB_2 y X_2 a eNB_1 o hacerlo ambos a través de un mismo eNB. Los valores de los KPIs serían los mismos. Tanto la señalización como los datos media de las llamadas del operador, atraviesan la red externa EXT_{net} en sentido eNB_1 <-> MAIN <-> eNB_2, eNB_1 <-> MAIN <-> eNB_1 o eNB_2 <-> MAIN <-> eNB_2. El retardo debido a T_{EXTnet} será el mismo en los tres casos.

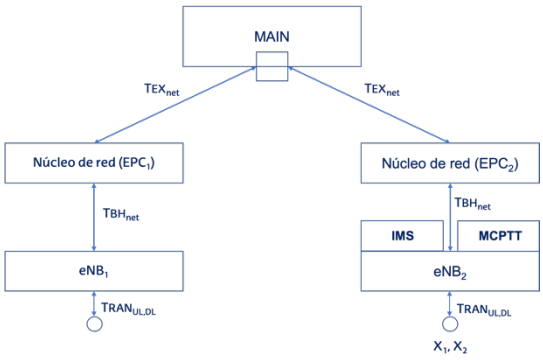


Fig. 5. OPCION 2 - MCPTT desplegado en el EDGE. Usuarios en el EDGE

- Opción 2: El servicio MCPTT se ha desplegado en el EDGE (eNB_2), Fig. 5. Los usuarios X_1 y X_2 acceden al servicio desplegado en el extremo. En esta opción se considera que todos los usuarios están concentrados en la misma celda y se conectan a través de eNB_2. Tanto la señalización como los datos media de las llamadas del operador, no atraviesan EXT_{net}, el sentido es eNB_2 <-> eNB_2 para los clientes X_1 y X_2. El servicio MCPTT se podría haber desplegado en el EDGE (eNB_1) y conectado los usuarios X1 y X2 al eNB_1. Se habrían obtenido los mismos valores para los KPIs.

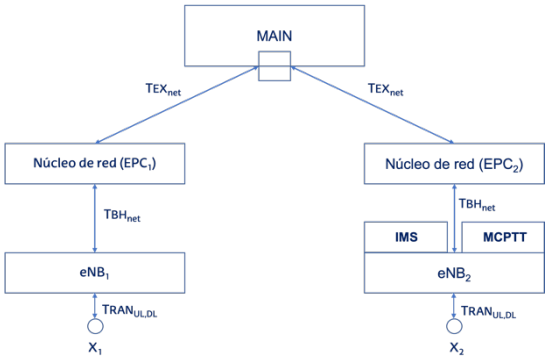


Fig. 6. OPCION 3: MCPTT desplegado en el EDGE del Operador. Usuarios distribuidos

- Opción 3 (Escenario Mixto): El servicio MCPTT está desplegado en el EDGE, Fig. 6. Gracias al simulador es posible simular qué ocurre con los KPIs cuando no todos los usuarios en llamada están conectados al nodo externo de red eNB$_2$ (EDGE) donde se ha desplegado el servicio.
- Una hipotética Opción 4 consistiría en desplegar el servicio MCPTT en el EDGE (eNB1), manteniendo los usuarios X1 y X2 como en la opción 3 pero no se ha representado explícitamente ya que el supuesto sería el mismo: unos usuarios tienen el servicio desplegado cerca de ellos y otros tienen que atravesar el MAIN para acceder a ellos.

C. Herramientas del simulador

Gracias al empleo de los *potenciómetros* de la caja de herramientas, denominada "Call Distribution in a Mix Scenario" Fig. 7, se puede repartir los usuarios en llamada entre eNB$_1$ y eNB$_2$.

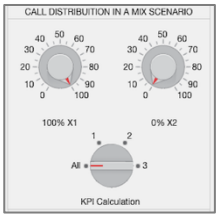


Fig. 7. Panel de Distribución de llamadas en la Opción 3

Tal y como están ajustados los potenciómetros en la figura anterior el 20 % de los usuarios totales en llamada son X$_1$ y el 80 % son X$_2$. Eso significa que para la opción 1, tantos los usuarios X$_1$ y X$_2$ tienen que subir hasta el MAIN donde esta desplegado el SIP CORE y el servicio MCPTT. Por tanto, el % de usuarios X$_1$ y X$_2$ resulta indiferente. TEXT$_{net}$, TEPC$_{net}$, TRAN$_{UL}$ y TRAN$_{DL}$ es el mismo en ambos casos. En la opción 2, los usuarios X$_1$ y X$_2$ acceden al MCPTT y SIP CORE desplegado en el extremo, mejorando los indicadores de servicio KPIs respecto de la opción 1 pero el % de usuarios X$_1$, X$_2$ de nuevo no afecta. Sin embargo, la opción 3, denominada mixta, la ubicación de los usuarios sí que va a afectar al rendimiento final del servicio. El servicio MCPTT y el SIP CORE están desplegados en el extremo, pero el impacto varía de forma sustancial si hay que atravesar el MAIN cuando un X$_1$ establece o participa una llamada frente a que las llamadas se hacen entre usuarios X$_2$.

El selector "**KPI Calculation**" permite seleccionar qué KPIs se quiere calcular, pudiendo analizarlos individual o simultáneamente los tres a la vez. Una vez configurado el simulador en el modo "**Auto Update**", cada vez que se desplaza uno de los dos potenciómetros se obtendrá directamente los KPIs de forma gráfica y alfanumérica.

Por defecto, el cálculo de los KPIs se realiza en base a los valores de los ficheros que contienen los tiempos de proceso o de red (latencia) medidos en cada uno de los componentes que intervienen en la llamada. Sin embargo, para hacer más dinámica esta simulación es posible variar Fig. 8 en - 50 y + 50 % el valor de estos tiempos.

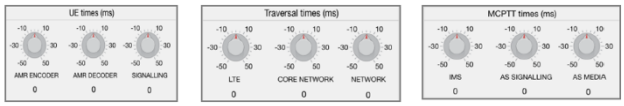


Fig. 8. Panel de ajuste tiempos por componente

En la parte derecha de la pantalla se puede ver en tiempo real cuál es el percentil 25, 50, 75 o 90 de los valores de entrada, pudiendo simular tantos escenarios como combinaciones de tiempos se esté interesado en analizar. Es posible simular, por ejemplo, si se produce una congestión en la RAN o si aumentan los tiempos de procesado de los servidores debidos al aumento de capacidad provocada por aumento de recursos debido a la emergencia.

En el gráfico y en el panel de datos se visualizan los KPIs seleccionados a través del selector "KPI Calculation". El formato del gráfico puede ser: "Line Graph" y "Bar Graph" y se selecciona en el desplegable "Graph Type".

La presentación de los resultados está dividida en dos zonas:

- **Panel de datos:** Se ubica en la parte inferior y es donde se representan los valores en tiempo real de cada uno de los KPIs seleccionados.

OP	KPI 1	KPI 2	KPI 3	DESCRIPTION
1	332.3089	543.6839	276.5728	MCPTT 100% Main Cloud
2	310.8487	500.2589	254.6101	MCPTT 100% Edge Cloud
3	345.2167	528.5210	270.3711	MCPTT Mix(80% In)+(20% Out)

Fig. 9. Panel alfanumérico de resultados

- **Panel de gráficos:** Se ubica sobre el panel de datos. Si el tipo de gráfico seleccionado es "Line Graph" se abrirá una solapa por cada uno de los KPIs seleccionados. Si el gráfico seleccionado es "Bar Graph" se abrirá una solapa por cada percentil; 0.95, 0.80, 0.75, 0.50, 0,25 y 0.20 que estará relacionado con el valor instantáneo de cada KPI para cada % de llamadas. De esta forma, se puede conocer si el 95% de las llamadas no cumplen el KPI, si lo hacen al 80%...20%. Esto va a permitir relacionar la distribución de los usuarios configurada en el escenario con el porcentaje de llamadas que cumplen cada KPI.

Tanto en la representación grafía como alfanumérica se representan las tres opciones de despliegues que contempla este simulador:

- **La opción 1:** Representada con el color azul.

- **La opción 2:** Representada con el color rojo.
- **La opción 3:** Representada con el color fucsia.

III. ANÁLISIS DE INDICADORES

A continuación, se van a analizar los indicadores de servicio KPI_1, KPI_2 y KPI_3 para cada una de las opciones de despliegue, modificando para cada opción el porcentaje de usuarios X_1 y X_2 del conjunto de usuarios X conectados. Los tiempos de procesado / latencia de los componentes de la comunicación no han sido modificados por los *potenciómetros* disponibles. Se emplean los tiempos sin ajuste. Se puede observar como todos los potenciómetros indican el valor 0, debajo de cada uno de ellos.

Por lo indicado anteriormente, en la sección II.B. Escenarios de Despliegue, se va a analizar cuándo el despliegue del servicio en el extremo de la red mejora los indicadores de servicio KPIs y comprobar cómo la ubicación de los usuarios que participan en la llamada es determinante para la toma de esa decisión. Para lo cual, se consideran diferentes distribuciones de usuarios.

Los tres gráficos Fig. 10, Fig. 11 y Fig. 12 que se presentan a continuación se han generado seleccionando la opción: Bar Graph. Se ha elegido esta opción ya que se pueden analizar de forma simultánea los valores de los tres indicadores, en las tres opciones de despliegue. Los valores que se muestran corresponden al 95 % de las llamadas analizadas y coincide con los requerimientos definidos por el 3GPP para definir el número de llamadas que deben cumplir con los requerimientos de latencia definidos para para indicador de servicio KPI_1, KPI_2 y KPI_3.

A. Distribución usuarios 100% X_1 / 0% X_2.

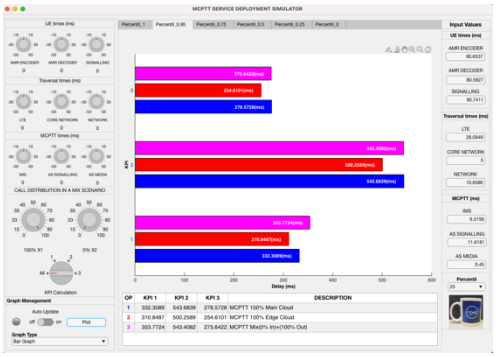


Fig. 10. Bar Graph – 100% X1 / 0% X2

Los valores que se han obtenido se representan en la tabla adjunta.

Tabla V
DISTRIBUCIÓN 100% X_1 y 0% X_2

Opciones	KPI 1 (ms)	KPI 2 (ms)	KPI 3 (ms)
1	332,3089	543,6839	276,5728
2	310,8487	500,2589	254,6101
3	353,7724	543,4082	275,6422

En este supuesto, la opción de despliegue que obtiene mejores indicadores de servicio para el 95% de las llamadas es el despliegue del SIP CORE y MCPTT en el MAIN. Con esta distribución de usuarios no es por tanto recomendable desplegar en el extremo de la red.

B. Distribución usuarios 10% X_1 / 90% X_2.

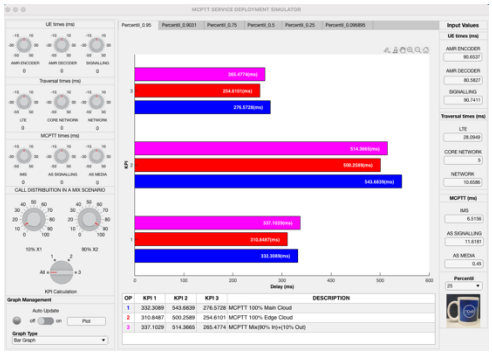


Fig. 11. Bar Graph – 10% X_1 / 90% X_2

Los valores que se han obtenido se representan en la tabla adjunta.

Tabla VI
DISTRIBUCIÓN 10% X1 y 90% X2

Opciones	KPI 1 (ms)	KPI 2 (ms)	KPI 3 (ms)
1	332,3089	543,6839	276,5728
2	310,8487	500,2589	254,6101
3	337,1029	514,3665	265,4774

En este supuesto, la opción de despliegue que obtiene mejores indicadores de servicio para el 95% de las llamadas es el despliegue del SIP CORE y MCPTT en el MAIN. El 10 % de usuarios X_1 penaliza mucho los indicadores de servicio para el 95% de las llamadas. Para el 75% de las llamadas se obtienen los siguientes valores.

Tabla VI
DISTRIBUCIÓN 10% X_1 y 90% X_2 (75% llamadas)

Opciones	KPI 1 (ms)	KPI 2 (ms)	KPI 3 (ms)
1	332,3089	543,6839	276,5728
2	310,8487	500,2589	254,6101
3	304,9674	489,3552	251,3923

Sin embargo, el requerimiento definido por el 3GPP es de 95% en lugar de 75%.

C. Distribución usuarios 5% X1 / 95% X_2.

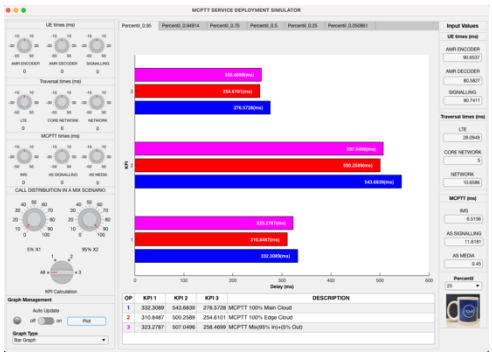


Fig. 12. Bar Graph – 5% X_1 / 95% X_2

Los valores que se han obtenido se representan en la tabla adjunta.

Tabla VII
DISTRIBUCIÓN 5% X_1 y 95% X_2

Opciones	KPI 1 (ms)	KPI 2 (ms)	KPI 3 (ms)
1	332,3089	543,6839	276,5728
2	310,8487	500,2589	254,6101
3	323,2787	507,0496	258,4699

En este supuesto, la opción de despliegue que obtiene mejores indicadores de servicio para el 95% de las llamadas es el despliegue del SIP CORE y MCPTT en el EDGE.

IV. CONCLUSIONES

Gracias al simulador desarrollado y el análisis de las contribuciones de los diferentes componentes se ha podido validar a través de datos reales el impacto del despliegue del servicio MCPTT en el extremo. De forma generalizada, una emergencia se localiza en un área geográfica concreta, eso implica que los recursos técnicos y por tanto las comunicaciones de emergencia entre ellos van a estar muy concentrados en un extremo de la red. En este simulador, se presenta un modelo de red simplificado en cuanto a estaciones base, pero suficiente para poder analizar el impacto que supone la localización de los recursos de emergencia en la red.

En el simulador se analizan las principales opciones de despliegue que se puede adoptar por parte del operador de red para garantizar el cumplimiento de los indicadores de servicio para el 95% de las llamadas.

El despliegue de los servicios en el extremo de la red permite acercar los servicios a los usuarios y por norma general la latencia debería disminuir. Sin embargo, en una llamada MCPTT tanto privada 1:1 como de grupo 1:n interviene más de un usuario. El impacto por tanto es mayor en las llamadas a grupo, donde los integrantes de este pueden estar muy distribuidos en la red.

En los datos presentados en este artículo se pone de manifiesto la importancia de conocer la ubicación de los usuarios en la red de cara a desplegar el servicio en la celda en donde se esté produciendo la mayor concentración de usuarios en la llamada.

Cualquier usuario que esté fuera de la celda en la que se ha desplegado el servicio puede afectar de forma más que significativa al rendimiento del servicio del conjunto de llamadas considerando el retardo adicional.

Si nos centramos en los datos obtenidos, solo el 5% de los usuarios en llamada podrían encontrase en otra ubicación. Esto es un dato significativo, ya que, de forma general, este grupo siempre esta supervisado por un puesto de mando que no se encuentra en la zona de emergencia y es el que en primer termino coordinará la intervención de ese grupo.

Según avance la emergencia, lo habitual será desplazar un puesto de mando avanzado a la zona donde se ha producido la emergencia y coordinar de forma local al grupo que está participando en la misma.

Gracias al simulador podremos modificar los valores de los diferentes tiempos de proceso y latencia que participan en la llamada mediante la manipulación de los potenciómetros correspondientes. Pudiendo aumentar o disminuir en un 50% el valor de componente.

Contrariamente a lo que se podría pensar, el despliegue del servicio MCPTT y el SIP CORE en el MAIN, es la opción (Opción 1) que se debería implementar si no se tiene información sobre la ubicación de los usuarios MCPTT en llamada (privada o de grupo).

Los resultados obtenidos confirman la mejora que supone en los KPIs de los usuarios en una llamada privada y grupo MCPTT que se encuentran conectados al eNB donde se ha desplegado el servicio MCPTT y el aumento de los KPI_S para aquellos usuarios conectados a otro eNB.

En una llamada a grupo MCPTT, el aumento del KPI_3 puede provocar que un usuario no reciba la información correctamente. Además, todo este aumento de tráfico media puede provocar un retardo en la señalización del canal que impediría a estos poder solicitar el canal para hablar o confirmar la recepción de la información. Siempre y cuando no se encuentren separados ambos tráficos.

Podría plantearse la posibilidad de desplegar, el servicio MCPTT en el MAIN y en el EDGE. No se resolvería el problema, ya que una misma llamada solo puede ser gestionada por uno de ellos (MAIN o EDGE) dando lugar de nuevo a las tres opciones de despliegue planteados. Sin embargo, aumentaría la complejidad de la arquitectura en la que habría que incluir un proxy para distribuir las llamadas o un balanceador de cargas para distribuirlas.

Bibliografía

[1] NPSTC,, "Push to Talk (PTT) over LTE Public Safety Requirements report", NPSTC.

[2] O. S. Ramon Ferrús, "Future Mobile Broadband PPDR Communications Systems, in Mobile Broadband Communications for Public Safety: The Road Ahead Through LTE Technology", Wiley, pp. 81-125, August 2015.

[3] NPSTC, “Defining Public Safety Grade Systems and Facilities, Final Report", NPSTC, 22/5/2014.

[4] 3GPP TS 22.179, , "Technical Specification Group Services and System Aspects; Mission Critical Push to Talk (MCPTT) over LTE; Stage 1”, V14.3.0 (2016-12): 3rd Generation Partnership Project (3GPP), December 2016.

[5] Aitor Sanchoyerto, R. Solozabal, B. Blanco and F. , "Analysis of the Impact of the Evolution Toward 5G Architectures on Mission Critical Push-to-Talk Services", in IEEE Access: vol. 7, pp. 115052-115061, 2019.

[6] Aitor Sanchoyerto Martinez, "Análisis del despliegue de comunicaciones de misión critica sobre redes 4G y 5G", https://addi.ehu.es/handle/10810/52175: Bilbao (Spain), 2021.

[7] D. Axiotis, D. Xenikos , “UDP Performance Measurements over TETRA IP”, n proceedings of VTC: Spring, 2007.

[8] 3. T. 23.379, "Functional architecture and information flows to support Mission Critical Push To Talk (MCPTT); Stage 2 (Release 14)", version 14.5.0: 3GPP, April 2018.

[9] 3. S. W. Group, "Mission critical applications", Available online at http://www.3gpp.org/specifications!groups/sa-plenary/sa6-missioncritical-applications, March 2015..

[10] C.C.B.G., "The Strategic Case for Mission Critical Mobile Broadband - A review of the future needs of of critical communications the users", Cambridgeshire, UK: TETRA and Critical Communications Association (TCCA), December 2013.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Aplicación basada en Blockchain para la Emisión y Validación de Certificados Académicos

Joan Amengual Mesquida, M. Magdalena Payeras Capellà, Macià Mut Puigserver, Llorenç Huguet Rotger
Departament de Ciències Matemàtiques i Informàtica,
Universitat de les Illes Balears,
Carretera de Valldemossa, Km. 7,5 07122 Palma.
jamengual150899@gmail.com, mpayeras@uib.cat, macia.mut@uib.cat, l.huguet@uib.cat

En los últimos años las credenciales académicas se han considerado las acreditaciones que permiten validar los conocimientos de las personas y así se ha visto reflejado en el número de nuevos matriculados y egresados por las universidades que ha aumentado de manera exponencial. Esta tendencia en el sector académico ha generado muchas ventajas para nuestra sociedad. Sin embargo, el elevado valor de los certificados ha provocado la falsificación de estos. El objetivo de este proyecto es proporcionar un protocolo para la emisión y la validación de títulos universitarios, mediante el cual se resuelva el fraude de dichos títulos. El protocolo resultante se basa en el uso de la tecnología blockchain que proporciona una estructura de datos pública, descentralizada, abierta e inmutable.

***Palabras Clave*—Blockchain, smart contract, seguridad, privacidad, certificados académicos, aplicación.**

I. INTRODUCCIÓN

La acreditación de los estudios universitarios se realiza mediante títulos o certificados académicos. Los certificados académicos son muy apreciados porque sirven de indicador del capital humano de sus titulares [1]. El capital humano se refiere a las habilidades, competencias, conocimientos y aptitudes alcanzadas a través de la educación [2]. Los títulos académicos son especialmente importantes en situaciones de empleo, ya que sirven como garantía no solo de los conocimientos, la experiencia y las aptitudes de sus titulares, sino también de sus capacidades, su fiabilidad y su dedicación.

Como los certificados universitarios son tan valiosos, las personas falsean sus calificaciones académicas presentando certificados falsos. En Estados Unidos hay 2 millones de certificados de grado falsos en circulación y 300 universidades no autorizadas en funcionamiento [3]. Estados Unidos tiene el mayor número de instituciones académicas falsas en el mundo, seguido por el Reino Unido, que cuenta con unos 270 institutos falsos [4]. Chiyevo et al. explica en [5] que los títulos académicos se consideran auténticos cuando los confiere una universidad legalmente autorizada para otorgar dichos certificados.

En el sistema educativo actual es posible realizar acciones fraudulentas por parte de diferentes actores. En particular, estas acciones pueden ser realizadas por:

- Los profesores que son aquellos que evalúan el trabajo realizado por los estudiantes.
- El personal de administración de la universidad que realiza la comprobación de la finalización de los estudios y emite el correspondiente certificado académico.
- Los estudiantes que son los individuos que presentan el certificado emitido por la universidad al lugar de trabajo donde deseen acceder.

Entonces, este conjunto de actores puede realizar acciones fraudulentas con las que se declaran que un individuo tiene unos conocimientos de los que carece.

Normalmente, las universidades hacen entrega de los certificados académicos en formato físico, es decir, mediante un papel que certifica que el estudiante ha cursado satisfactoriamente unos estudios en dicha universidad y ha obtenido el título universitario de manera válida. Seguidamente, los estudiantes pueden presentar los certificados académicos en formato físico en su lugar de trabajo mediante el uso de copias compulsadas o mediante el certificado entregado por la misma universidad.

En cambio, hay universidades que entregan los títulos universitarios en formato digital. Los departamentos de recursos humanos tienen que confiar en las bases de datos

de las universidades donde se almacenan los datos de los certificados de los estudiantes. Estas bases de datos tradicionales no son a prueba de manipulaciones, son propensas a ser comprometidas o pueden ser modificadas por cualquier funcionario interno. Y no es posible que un empleador detecte la manipulación en la base de datos de la universidad. Por lo general, estas bases de datos tradicionales se encuentran en un servidor centralizado, no son transparentes y solo pueden acceder a ellas los administradores de la base de datos. No hay forma de rastrear o verificar los datos del certificado directamente en la base de datos de la universidad. Por lo tanto, los usuarios tienen que confiar ciegamente en el sistema de gestión de datos de estudio de la universidad para la legitimidad del certificado [6].

Para la resolución de esta problemática se puede hacer uso de la tecnología blockchain. Según un informe del servicio de ciencia y conocimiento de la Comisión Europea [7] blockchain es un área de creciente interés para muchas industrias y universidades a nivel mundial. Blockchain es una tecnología transversal, intersectorial y disruptiva que, según las previsiones, impulsará el crecimiento de la economía mundial durante las próximas décadas.

La tecnología blockchain ofrece características como el almacenamiento de datos decentralizado, transparente y a prueba de manipulaciones. Puede utilizarse para resolver problemas como la falta de confianza, el fraude, el alto coste de las transacciones, la compartición, la privacidad y la evaluación de la fiabilidad de un actor potencial en una transacción. Por lo tanto, blockchain es una tecnología prometedora para prevenir las actividades fraudulentas en nuestro actual sistema de certificados académicos [6].

La tecnología blockchain es ideal como nueva infraestructura para asegurar, compartir y verificar los logros del aprendizaje. En el caso de las certificaciones académicas, una cadena de bloques puede mantener una lista de emisores y receptores de cada certificado académico, junto con la firma del documento (*hash*) en una base de datos pública (la cadena de bloques) que se almacena de forma idéntica en miles de ordenadores.

Los certificados académicos digitales asegurados en una cadena de bloques tienen ventajas significativas sobre los certificados digitales ”tradicionales”:

- No pueden ser falsificados, ya que es posible verificar con certeza que el certificado fue originalmente emitido y recibido por las personas indicadas en el certificado.
- La verificación del certificado puede ser realizada por cualquier persona que tenga acceso a la cadena de bloques, con software de código abierto fácilmente disponible no hay necesidad de ninguna parte intermediaria.
- Al no ser necesaria ninguna parte intermediaria para validar el certificado, este puede seguir siendo validado incluso si la organización que lo emitió ya no existe o ya no tiene acceso al registro emitido.
- El registro de los certificados emitidos y recibidos en una cadena de bloques solo puede destruirse si se eliminan todas las copias en todos los ordenadores del mundo que albergan el software.

Actualmente existen varios sistemas que pretenden combatir la falsificación de certificados académicos, en particular se destacan los siguientes:

- Blockcerts [8]
- EduCTX [9]
- Blockchain for Education [10]

En la tabla I se comparan las características de cada uno de estos sistemas con sus ventajas e inconvenientes.

Cabe mencionar el proyecto español que pretende impulsar una prueba de concepto generando una red Blockchain, denominada Blue (BLockchain Universidades Españolas) para comenzar a desarrollar y desplegando servicios en ella. Como primer servicio desarrollado sobre esta nueva red se propone la emisión de certificados.

II. CONTRIBUCIÓN DEL PROYECTO

Las contribuciones y mejoras más destacables en las que contribuye el proyecto son las siguientes:

- Erradicar la inundación de títulos universitarios en el mercado laboral por las fábricas de certificados y así evitar que personas sin la titulación requerida lleguen a optar a trabajos mediante títulos universitarios falsificados teniendo ventaja sobre las personas que han cursado los estudios correctamente.
- Proporcionar un sistema de emisión de títulos por parte de entidades autorizadas con validación universal.
- Diseñar e implementar un protocolo donde participan todas las entidades involucradas en el proceso de emisión y validación de títulos universitarios mediante la tecnología blockchain.

III. PROTOCOLO DE EMISIÓN Y VALIDACIÓN DE TÍTULOS UNIVERSITARIOS

El protocolo para la emisión y la validación de títulos universitarios se ha dividido en tres bloques:

1) Validación de las universidades como autoridades certificadoras
2) Publicación de certificados académicos en la blockchain
3) Validación de certificados académicos

Algunos de los elementos que han posibilitado el desarrollo del protocolo han sido:

Tabla I
COMPARACIÓN ENTRE LOS SISTEMAS EXISTENTES BASADOS EN BLOCKCHAIN.

	Sistemas basados en blockchain		
	Blockcerts [8]	**EduCTX** [9]	**Blockchain for Education** [10]
Tipo de certificado o logro	Cualquier tipo de certificado académico y no académico	Obtención de créditos por logros académicos completados, como ECTS	Certificados de educación superior o instituciones acreditadas
Emisor del certificado	Cualquier tipo de instituto; académico y no académico; educación formal e informal	Instituciones de educación superior que siguen los estándares del ECTS y se han unido a la red	Instituciones acreditadas
Plataforma blockchain	Bitcoin, Ethereum	Ark	Ethereum
Uso de Smart Contract	No	No	Sí
Accesibilidad en la blockchain	Pública	Privada / Red de consenso	Pública
Protocolo de consenso	Proof of Work (PoW)	Delegated Proof of Stake (DPoS)	Proof of Work (PoW)
Uso de IPFS	No	No	Sí
Datos guardados en la blockchain	Hashes de los certificados (en un lote) Se almacena la raíz de Merkle	Tokens ECTX, identificación del curso, identificación del emisor (Instituto) y la identificación del receptor (estudiante)	Hash de los certificados, claves públicas de las autoridades de certificación, hash de la dirección IPFS de la información del perfil de las autoridades de certificación
Ventajas	- Estándar abierto - Cumplimiento de los criterios de auto-soberanía digital - Reducción del coste de transacción por certificado mediante el uso de árboles de Merkle	- Registra la información de cada curso, no solo del certificado final	- Solo las universidades acreditadas pueden emitir certificados - Revocación y renovación de los certificados - Los estudiantes pueden decidir aquello que pueden hacer los empleadores con sus certificados (leer o verificar)
Inconvenientes	- Vulnerable a los ataques de suplantación de identidad de emisores de los certificados	- El proceso no está totalmente automatizado y existe la posibilidad de cometer errores al transferir una cantidad incorrecta de tokens - El estudiante no puede seleccionar que registros compartir con el empleador	- Costes de las transacciones por cada certificado - Necesidad de verificar la autoridad de certificación previamente

- IPFS (InterPlanetary File System): Sistema de archivos distribuido.
- Smart Contract: Programa informático que facilita, asegura, hace cumplir y ejecuta acuerdos registrados entre dos o más partes.

Durante la explicación del protocolo se hará referencia a dos autoridades: autoridad de acreditación y autoridad certificadora.

- Autoridad de acreditación. Tiene como función validar a las universidades para que puedan convertirse en autoridades certificadoras y poder así publicar certificados académicos.
- Autoridad certificadora. Son aquellas universidades validadas por la autoridad de acreditación que disponen de los permisos para publicar certificados académicos en la cadena de bloques.

Estas son las autoridades principales del protocolo y forman una estructura jerárquica entre sí.

A. Validación de las universidades como autoridades certificadoras

La primera fase del protocolo consiste en validar a las universidades para que puedan convertirse en autoridades certificadoras, y de esta forma que dispongan de los permisos necesarios para publicar certificados académicos en la cadena de bloques.

En consecuencia, se debe definir una autoridad de acreditación. En este caso la autoridad de acreditación es un organismo regulador que permite decidir que universidades tienen derecho a publicar certificados. En este protocolo, se ha establecido como autoridad de acreditación al Ministerio de Universidades.

1) Alta de una universidad como autoridad certificadora: Cuando una universidad desee publicar certificados académicos a la cadena de bloques deberá comunicarlo al Ministerio. Seguidamente se va a iniciar un proceso de verificación de la universidad para decidir si puede convertirse en autoridad certificadora. Para acreditar a la universidad como autoridad certificadora en caso de haber sido verificada previamente, el Ministerio deberá realizar el siguiente procedimiento:

1) El Ministerio deberá disponer de la dirección o de las direcciones de la wallet de la universidad desde las que se van a publicar los certificados a la cadena de bloques. Esta dirección o direcciones deberán ser comunicadas entre la universidad y el Ministerio.

2) El Ministerio deberá identificar y almacenar las direcciones aportadas por las universidades en sus bases de datos. De esta manera, el Ministerio tendrá constancia en todo momento de cuales son las direcciones utilizadas por cada universidad para la

publicación de certificados.

3) Para dar de alta a una universidad como autoridad certificadora el Ministerio deberá introducir la dirección de la wallet de la universidad en un campo específico de la interficie de la aplicación.

   A continuación, se va a detallar el flujo de comunicación entre el Ministerio y el Smart Contract para validar una dirección de una universidad. Si la universidad desea aportar más direcciones desde las que publicar certificados el proceso deberá repetirse tantas veces como direcciones se deseen acreditar:
   - La dirección de la wallet de la universidad pasará a registrarse en la lista de las autoridades certificadoras del Smart Contract.
   - La dirección de la wallet de la universidad pasará a tener un estado válido a partir de este momento, el cual podrá cambiar si el Ministerio lo desea en cualquier momento, eliminando así su función de autoridad certificadora.

4) El Ministerio podrá notificar a la universidad que ya pertenece al conjunto de autoridades certificadoras. A partir de este momento, la universidad ya tendrá el derecho a publicar títulos universitarios en la cadena de bloques.

2) Baja de una universidad como autoridad certificadora: Es posible que después de un tiempo la universidad desee darse de baja como autoridad certificadora, y por tanto dejar de tener el derecho a publicar certificados a la cadena de bloques. Además, es posible que el Ministerio considere por alguna razón que dicha universidad no debe poder seguir publicando certificados. Por lo que se ha considerado la funcionalidad para dar de baja a una universidad como autoridad certificadora. Tal y como se puede visualizar en la Figura 1 el Ministerio deberá realizar el siguiente procedimiento:

1) El Ministerio deberá disponer de la dirección o de las direcciones utilizadas por la universidad para realizar la publicación de certificados.

2) El Ministerio deberá introducir la dirección de la wallet de la universidad en un campo específico de la web. A continuación, se va a detallar el flujo de comunicación entre el Ministerio y el Smart Contract, en caso de que la universidad disponga de varias direcciones deberá repetirse el proceso tantas veces como direcciones se deseen desacreditar:
   a) La dirección de la wallet de la universidad pasará a registrarse en la lista de las universidades dadas de baja del Smart Contract.
   b) La dirección de la wallet de la universidad pasará a tener un estado no válido a partir de este momento, es decir, la universidad no podrá publicar más certificados académicos a la cadena de bloques.

3) El Ministerio podrá notificar a la universidad que ya no pertenece al conjunto de autoridades certificadoras. Y por tanto, ya no se le permite publicar certificados a la cadena de bloques.

En la Figura 1 los recuadros identificados con (A) enmarcan la comunicación previa entre la universidad y el Ministerio en los procesos de dar de alta y de baja a las universidades, los recuadros identificados con (B) enmarcan la comunicación entre el Ministerio y el Smart Contract y los recuadros identificados con (C) enmarcan las notificaciones de los procesos de dar de alta y de baja del Ministerio a la universidad.

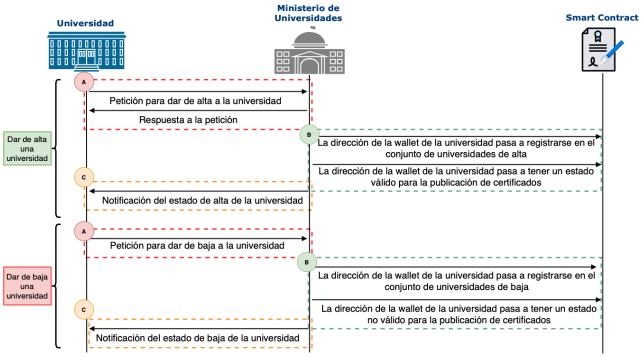


Fig. 1. Diagrama del proceso de dar de alta y de baja a universidades como autoridades certificadoras.

B. Publicación de certificados en la blockchain

Una vez la universidad se ha acreditado como autoridad certificadora tendrá el derecho a la publicación de certificados en la cadena de bloques. Este proceso se ha detallado mediante la Figura 2. En primer lugar, la universidad debe disponer de un conjunto de datos del estudiante necesarios para la identificación de este. La comunicación entre el estudiante y la universidad será realizada off-chain, es decir van a comunicarse por el canal de comunicación que ellos consideren adecuado.

En el diagrama de la Figura 2 se hace referencia a la petición de un código secreto como uno de los datos personales requeridos. Una de las desventajas del protocolo EduCTX recae en que el estudiante no puede seleccionar los registros a compartir con el empleador. Por tanto, mediante el valor del código secreto se pretende solucionar esta problemática.

Este campo va a ofrecer a los estudiantes la posibilidad de presentar los certificados de manera individual o en lotes. El estudiante podrá presentar un certificado aportando el código secreto con el que se enlazó en su publicación o podrá presentar un conjunto de certificados enlazados a un mismo código secreto.

El uso de este código secreto permite que este proyecto escale en un y permita publicar certificados de distintos ámbitos. Por ejemplo, certificados que acrediten un nivel de idiomas. El estudiante tendrá la posibilidad de enlazar certificados con diferentes códigos, y presentarlos de manera independiente o en lotes. La longitud del código secreto se ha establecido en un mínimo de 6 dígitos para incrementar la seguridad del proceso de validación.

Una vez la universidad disponga de los datos personales del estudiante requeridos, deberá publicar el PDF del certificado a la IPFS, y esto posibilitará visualizar el certificado en formato PDF una vez se requiera su validación. La IPFS va a proporcionar un hash del certificado con el que se podrá identificar de manera única al certificado. En este protocolo el hash proporcionado por IPFS se va a nombrar código IPFS.

En este protocolo los certificados requieren de la existencia de una referencia electrónica, similar al uso del CSV (Código Seguro de Verificación) usado por la Agencia Tributaria para identificar documentos electrónicos de forma única. Mediante la referencia electrónica se podrá identificar al certificado y se requerirá para la publicación del certificado a la cadena de bloques.

Cuando la universidad disponga de los datos del estudiante y del código IPFS CID (Content Identifier) del certificado académico ya podrá publicar el título a la cadena de bloques. Para su publicación la universidad deberá completar un conjunto de campos documentados en la Tabla II.

Tabla II
CONJUNTO DE LOS CAMPOS A COMPLETAR PARA LA PUBLICACIÓN DE UN CERTIFICADO ACADÉMICO A LA CADENA DE BLOQUES.

Campo	**Ejemplo**
Referencia electrónica del certificado	2IYRM-M2XL5-TDB6E-XCNCN-7HQXF-QFYMH
Código IPFS (CID) del certificado	Qmei1VwBQ9o7ADcjPFYovJbWbq8yQeoWosuTn5kLnHMJRJ
Identificación del estudiante (NIF/NIE)	12345678X
Código secreto (mínimo de 6 digitos)	123456

Finalmente, la universidad podrá publicar el certificado académico a la cadena de bloques, y el Smart Contract realizará el siguiente procedimiento:

1) La referencia electrónica del certificado se va a enlazar con el código IPFS (CID) y con el hash de la identificación del estudiante (NIF/NIE). Así en el proceso de validación se podrá validar un certificado mediante su referencia electrónica y además visualizar su contenido en IPFS.

2) Se va a realizar el hash de la identificación del estudiante y del código secreto del estudiante de manera conjunta, y se va a enlazar este hash con el código IPFS (CID).
Es necesario destacar que el código IPFS no se va a cifrar, por tanto, será visible por cualquier usuario y en consecuencia, cualquier persona podrá visualizar los datos de los certificados. No obstante, no se consideran datos sensibles aquellos datos que se encuentran en un certificado académico y es preferible eliminar la carga del cifrado.
Este enlace entre el hash de la identificación del estudiante y el código secreto con el código IPFS de su título será necesario para poder visualizar el conjunto de certificados de los que dispone un estudiante ligados a un mismo código secreto.
Un usuario podrá visualizar el PDF de los certificados académicos en IPFS que se encuentren ligados a una identificación (NIF/NIE) y a un código secreto.

3) Finalmente, se va a almacenar el número de certificados publicados por cada universidad. Así, se podrá analizar el uso que se está dando al servicio por cada universidad.

En la Figura 2 el recuadro identificado con (A) enmarca la comunicación entre el estudiante y la universidad, el recuadro identificado con (B) enmarca la comuicación entre la universidad y la IPFS y el recuadro identificado con (C) enmarca la comunicación entre la universidad y el Smart Contract.

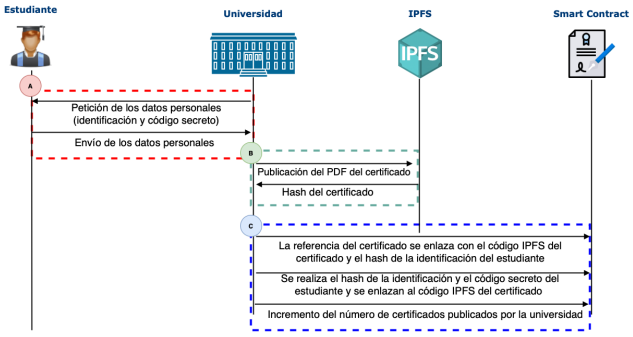


Fig. 2. Diagrama del proceso de publicación de certificados en la cadena de bloques por una universidad.

C. *Validación de certificados*

La validación de certificados es la última fase del protocolo. En este apartado, se va a detallar su proceso mediante la Figura 3, donde se describen las comunicaciones entre los actores implicados.

En primer lugar, cabe destacar que cualquier usuario tendrá la posibilidad de validar certificados académicos. En particular, se va a centrar el punto de interés de la validación de los certificados a los empleadores. La validación de los certificados se puede realizar de distintas formas:

- Validación mediante la referencia electrónica del certificado. Consiste en validar un certificado de manera independiente, es decir, el proceso de validación únicamente se enfoca en validar un certificado.
- Validación mediante el código secreto. Posibilita la validación de un conjunto de certificados en un mismo proceso de validación.

1) Validación mediante la referencia electrónica del certificado: El desarrollo para realizar la validación mediante la referencia electrónica va a permitir validar y visualizar un único certificado académico. El proceso a efectuar es el siguiente:

1) El empleador debe disponer de dos parámetros: la referencia electrónica del certificado y la identificación del estudiante (NIF/NIE). En vista de ello, el estudiante deberá aportar la información previamente para efectuar la validación del certificado.

2) Cuando el empleador introduzca los datos necesarios (referencia electrónica del certificado y la identificación del estudiante) el Smart Contract va a relacionar la referencia del certificado académico y la identificación del estudiante (NIF/ NIE) con el código IPFS (CID) del certificado. De esta manera, se podrá mostrar el código IPFS que identifica el PDF del certificado en IPFS.

3) Finalmente, el empleador dispondrá del código IPFS (CID) del certificado y podrá proceder a visualizar su contenido mediante el uso de IPFS.

Con este procedimiento el empleador puede validar un certificado académico y además visualizar su contenido con el uso de IPFS.

2) Validación mediante el código secreto: El desarrollo para realizar la validación mediante el código secreto va a permitir validar y visualizar varios certificados en un mismo proceso. Este procedimiento se ha detallado en la Figura 3. El proceso a efectuar es el siguiente:

1) El empleador debe disponer de dos parámetros: la identificación del estudiante (NIF/NIE) y el código secreto enlazado al conjunto de certificados que se deseen validar. Por tanto, el estudiante deberá aportar dicha información previamente para realizar la validación de los certificados.

2) Cuando el empleador introduzca los datos necesarios (identificación del estudiante y código secreto) el Smart Contract podrá relacionar la identificación del estudiante y el código secreto con los códigos IPFS de los certificados. Por tanto, se podrán mostrar los códigos IPFS que identifican al PDF de cada uno de los certificados en IPFS.

3) Finalmente, el empleador dispondrá de los códigos IPFS de los certificados y podrá proceder a visualizar su contenido mediante IPFS.

Con este procedimiento el empleador puede validar un certificado académico o un conjunto de estos. Además podrá visualizar su contenido con el uso de IPFS.

En la Figura 3 los recuadros identificados con (A) enmarcan la comunicación entre el empleador y la Smart Contract y los recuadros identificados con (B) enmarcan la comunicación entre el empleador y la IPFS.

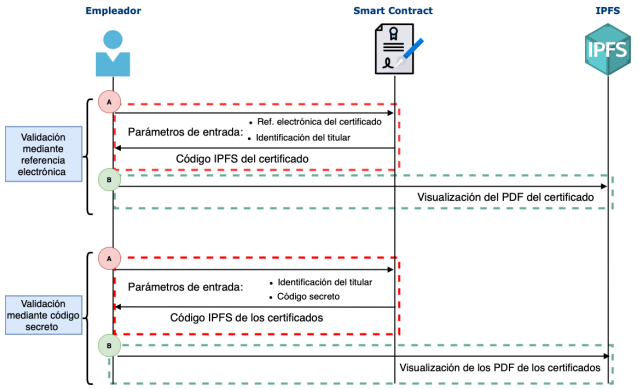


Fig. 3. Diagrama del proceso de validación de certificados mediante la referencia electrónica y mediante el código secreto

IV. ANÁLISIS DE PROPIEDADES

A. Análisis del protocolo

Los tipos de certificados o logros que contempla el protocolo desarrollado son puramente certificados académicos, específicamente títulos universitarios. Este protocolo se asemeja al protocolo Blockchain for Education, donde el punto de mira se centra en certificados de educación superior o de instituciones acreditadas.

Un punto que se ha considerado desde el primer momento en el desarrollo del protocolo ha sido el emisor del certificado académico y la validez de este. En sistemas como Blockcerts cualquier institución puede emitir certificados, esta es una característica del sistema que puede resultar positiva, ya que permite tener un amplio abanico de aplicaciones. No obstante, esta misma característica resulta ser negativa en el momento que se consideran certificados altamente valiosos como es el caso de los títulos universitarios. Así que en estos certificados es relevante destacar cuál ha sido el emisor para poder validar completamente el certificado académico. Por ello, en el protocolo desarrollado se ha definido una autoridad reguladora encargada de validar que instituciones tienen la potestad de publicar los certificados académicos sobre la cadena de bloques. Únicamente las universidades que se validen por esta autoridad reguladora serán las permitidas para la publicación de certificados académicos

en la cadena de bloques.

La plataforma blockchain elegida para la implementación del proyecto ha sido Ethereum. Los principales motivos de esta elección han sido los siguientes:

- Ethereum ha creado una plataforma de acceso global en la que se pueden ejecutar complejos contratos en red. Elimina totalmente la necesidad de que existan servicios proporcionados por terceros para su funcionamiento.
- Ethereum sirve de plataforma para otros productos o servicios, eso permite que se cree un ecosistema robusto que hace cada vez más fuerte a la plataforma. A medida que avancemos en el tiempo, cada vez habrá mejor información sobre Ethereum.
- Más allá de los fundadores de Ethereum, existen muchas compañías involucradas en su estudio y desarrollo, como Ethereum Enterprise Alliance o el equipo de Hyperledger. Después de Bitcoin es la blockchain con mayor apoyo de la comunidad empresarial.

Un aspecto a considerar en este proyecto ha sido el desarrollo de un Smart Contract que permita regular todas las operaciones a realizar. A diferencia de los protocolos Blockcerts y EduCTX donde no se usa ningún Smart Contract para su funcionamiento. Además, se ha tenido mucha consideración en los datos a almacenar en la cadena de bloques debido a que la publicación de información en la cadena de bloques tiene un coste.

En el sistema implementado se ha hecho uso de IPFS como herramienta de visualización de los certificados. De la misma manera que Blockchain for Education lo usa para disponer de información de las autoridades de certificación. En los otros sistemas, Blockcerts y EduCTX, no se hace uso de IPFS.

La información que almacena el sistema Blockcerts es la raíz de Merkle que permite la validación de un lote de certificados. EduCTX almacena Tokens ECTX y otros datos relevantes para la validación de los certificados y Blockchain for Education almacena el hash de los certificados, claves públicas de las autoridades de certificación y otros datos. En el protocolo desarrollado se ha decidido almacenar las identificaciones de los certificados como sus referencias electrónicas y sus códigos IPFS enlazados a los *hashes* de las identificaciones de los titulados.

En la tabla III se destacan las características del protocolo desarrollado.

Tabla III
CARACTERÍSTICAS DESTACADAS DEL PROTOCOLO IMPLEMENTADO EN ESTE PROYECTO.

	Protocolo diseñado
Tipo de certificado o logro	Títulos universitarios, no obstante, puede considerarse su uso para cualquier certificado académico
Emisor del certificado	Universidades acreditadas por una autoridad reguladora
Plataforma blockchain	Ethereum
Uso de Smart Contract	Sí
Accesibilidad en la blockchain	Pública
Protocolo de consenso	Proof of Work (PoW)
Uso de IPFS	Sí
Datos guardados en la blockchain	Hash de las referencias electrónicas de los certificados y códigos IPFS de los certificados enlazados a las identificaciones de los titulados

V. ANÁLISIS DE RENDIMIENTO

Uno de los factores claves a la hora de desarrollar tecnologías basadas en blockchain es el análisis de rendimiento debido a que en estas tecnologías el código no es ejecutado sobre la máquina local, sino sobre una red P2P distribuida. Esto significa que el coste computacional de ejecución de un contrato se traduce en un coste económico determinado y un cierto tiempo de espera. A continuación se valorarán estos dos parámetros en la implementación que se ha realizado del protocolo.

Es importante establecer en primer lugar, que estos parámetros se han estudiado de acuerdo a las pruebas establecidas sobre la red Rinkeby. Esta red no es más que una red privada de prueba para Ethereum, dedicada específicamente para desarrolladores. Es por ello que los resultados obtenidos en esta sección, especialmente los resultados temporales, deben considerarse orientativos y servir únicamente para establecer referencias.

A. Tiempo de espera

En primer lugar se evaluará el tiempo promedio que tarda en computarse cada una de las distintas funcionalidades del protocolo. Como ya se ha expuesto anteriormente, esto se realizará en la red de pruebas Rinkeby. Esta red tarda en promedio 15 segundos en publicar un nuevo bloque.

El tiempo de espera se ha calculado como la diferencia entre el tiempo de publicación del bloque correspondiente y la entrada de la transacción en la red *P2P*. Para ello se ha hecho uso de la herramienta etherscan. Adicionalmente, cada acción se ha realizado un total de 10 veces con el objetivo de proporcionar la media aritmética de estos resultados.

Tabla IV
TIEMPOS MEDIOS DE EJECUCIÓN DEL PROTOCOLO.

Función	**Tiempo (segundos)**
Creación del Smart Contract	2,3
Alta de una universidad	5,5
Baja de una universidad	5,7
Publicación de certificados	6,5

Tal y como se puede apreciar en la Tabla IV, los resultados temporales son prácticamente idénticos, puesto que el factor que determina el tiempo de espera es común a todas las funciones. Esto se debe a que las funciones que han sido medidas necesitan la publicación de un nuevo bloque para realizarse. Esto significa que el tiempo mostrado en la tabla se ve afectado por el intervalo de publicación de bloques sumado al tiempo de aceptación de la transacción, el cual va variando en función del tráfico de la red *P2P* y el precio de la transacción. Por lo tanto, no es posible medir con exactitud el tiempo de ejecución de cada uno de los métodos, ya que ninguno requiere un tiempo de ejecución tan elevado como para no encontrarse totalmente influenciado por los tiempos de aceptación y publicación.

B. Costes de Ejecución

Este apartado pretende determinar el coste de gas para cada una de las funciones del Smart Contract.

Tabla V
COSTES DE EJECUCIÓN DE LAS FUNCIONES DEL SMART CONTRACT.

Función	**Gas (weis)**	**USD (1Gwei)**	**USD (20Gwei)**
Creación del Smart Contract	1.367.823	2,49	49,77
Alta de una universidad	130.179	0,24	4,74
Baja de una universidad	107.712	0,20	3,92
Publicación de certificados	359.577	0,65	13,08

Las medidas mostradas en la Tabla V corresponden a los costes fijos de gas de cada uno de los métodos del contrato. Adicionalmente, se ha añadido a modo orientativo el precio en dólares americanos de cada una de las funcionalidades teniendo en cuenta el valor al cambio entre monedas a 18 de marzo de 2021. Es importante mencionar que estos costes se han realizado teniendo en cuenta un precio de 1 Gwei y 20 Gwei (valor máximo). La principal diferencia, además del coste total de la transacción, es el tiempo que tardará la transacción en ser aceptada por un nodo minero. Por tanto, como mayor sea el precio que se pague por una misma transacción, menor será el tiempo de aceptación de la transacción. No obstante, para la publicación de títulos universitarios la velocidad en la publicación de estos no es un factor relevante y por ello se puede tener un coste mucho menor.

Las funciones de dar de alta y dar de baja a una universidad como autoridad certificadora tienen un precio muy similar debido a que la estructura del código es prácticamente idéntica. Se puede considerar un precio realmente bajo dar de alta y de baja a una universidad por un precio de alrededor 0,20 dólares. Este gasto podría ser asumido por la autoridad de acreditación.

En la publicación de certificados a la cadena de bloques el precio aumenta frente a los dos anteriores funciones, esto se debe a que en la publicación de certificados se realizan un conjunto de cálculos que provocan un aumento considerable del gas de ejecución. No obstante, podemos considerar este precio realmente bajo frente a las ventajas de tener el certificado publicado en la cadena de bloques y considerando las altas tasas actuales para la expedición de los títulos académicos. Este gasto podría ser asumido por la autoridad certificadora o por el titulado.

VI. CONCLUSIONES

Actualmente, las empresas carecen de herramientas para validar los certificados académicos que se les presentan. Esto ha posibilitado que las fábricas de diplomas inunden de certificados falsos el mundo laboral.

Recientemente, a través del desarrollo de nuevas tecnologías como blockchain se han abierto nuevas posibilidades para la resolución de estos problemas, permitiendo que las empresas tengan a su disposición herramientas para validar los certificados académicos presentados de manera rápida y fiable.

En este documento se ha presentado un protocolo para la emisión y validación de títulos universitarios con blockchain. En este protocolo se ha dado mucha importancia al proceso de emisión de los certificados, evitando así que cualquier institución sin acreditación sea capaz de emitir certificados académicos. También, se ha desarrollado un proceso de validación adecuado a cualquier persona independientemente de sus conocimientos sobre la tecnología blockchain.

AGRADECIMIENTOS

El proyecto FeltiCHAIN (RTI2018-097763-B-I00) está financiado por: FEDER/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación, (MCI/AEI/FEDER, UE).

REFERENCIAS

[1] O. Ghazali and O. Saleh, "A Graduation Certificate Verification Model via Utilization of the Blockchain Technology", 2018.
[2] O. Saleh, O. Ghazali and M. E. Rana, "Blockchain based framework for educational certificates verification", vol. 7, 2020.
[3] Grolleau, Gilles, Lakhal, and Mzoughi, "An Introduction to the Economics of Fake Degrees", Journal of Economic Issues, vol. 42, pp. 673-693, 2018.
[4] E. Ben and R. Winch, "Diploma and Accreditation Mills: New Trends in Credential Abuse", 2011, [Online]. Available: https://www.esrcheck.com/file/Verifile-Accredibase_Diploma-Mills.pdf.
[5] Evelyn Garwe, "Qualification, Award and Recognition Fraud in Higher Education in Zimbabwe",vol. 5, pp. 119-135, Critical Studies in Education, 2015.
[6] Rakibul Hasan, "Potencial of Blockchain technology to solve fake diploma problem", 2019.
[7] G. Alex and C. Anthony, "Blockchain in Education", 2017.
[8] David García, "Diseño de una lógica de negocio en blockchain. Desarrollo y despliegue de Smart Contracts en una Blockchain", pp 15-15, 2020.
[9] M. Turkanović, M. Hölbl, K. Košič, M. Heričko and A. Kamišalić, "EduCTX: A Blockchain-Based Higher Education Credit Platform", vol. 6, pp. 5112-5127, 2018.
[10] W. Gräther, S. Kolvenbach, R. Ruland, J. Schütte , C. Torres and F. Wendland, "Blockchain for Education: Lifelong Learning Passport", 2018.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Protocolo Basado en Blockchain para la Gestión de Canales para Microcompras Equitativas

M. Magdalena Payeras Capellà, Miquel À. Cabot-Nadal, Macià Mut Puigserver
Departament de Ciències Matemàtiques i Informàtica, Universitat de les Illes Balears
Crta. de Valldemossa, Km, 7.5 07122, Palma
mpayeras@uib.cat, miquel.cabot@uib.cat, macia.mut@uib.cat
Jordi Castellà Roca
Departament d'Enginyeria Informàtica i Matemàtiques, UNESCO Chair in Data Privacy, Universitat Rovira i Virgili
Av. Països Catalans 26, E-43007 Tarragona
jordi.castella@urv.cat

El desarrollo de aplicaciones de comercio electrónico que requieren el pago de pequeñas cantidades de dinero para la compra de servicios o bienes presenta desafíos en los campos de la seguridad y la privacidad. Estos pagos se denominan micropagos y ofrecen un equilibrio entre los requisitos de eficiencia y seguridad para pagar artículos de bajo valor. Con la aparición de las criptomonedas nuevos protocolos de compra pueden proponerse, beneficiándose de las cualidades de estos sistemas de pago. En este artículo se presenta un protocolo de compra o intercambio de pago por producto o recibo utilizando canales de pago. El sistema presenta las particularidades de permitir pagos de poca cuantía y proporcionar un intercambio justo entre la criptomoneda y el bien o servicio deseado. Finalmente, el protocolo incluye operaciones para el tratamiento de los pagos recibidos, como la transferibilidad que convierte al sistema en *multihop* o la posibilidad de reembolso del dinero no gastado en el canal.

***Palabras Clave*—Micropago, microcompra, canal, equidad, blockchain**

I. INTRODUCCIÓN

El campo del comercio electrónico (e-commerce) evoluciona día a día introduciendo nuevas aplicaciones y servicios. Una de estas aplicaciones es el pago de pequeñas cantidades de dinero para adquirir bienes o servicios de bajo coste. Este tipo de pagos se denomina micropago y tienen requisitos funcionales y de seguridad únicos dentro del campo de los pagos electrónicos. Los micropagos se pueden aplicar fácilmente a la venta intangible de bienes como información (periódicos, reseñas de productos, servicios basados en la ubicación, etc.), obsequios virtuales o datos electrónicos (música, vídeos, etc.). Todos estos ejemplos involucran transacciones de bajo valor, por lo que el costo operativo debe ser lo más bajo posible para que sea rentable para comerciantes y clientes.

Por un lado, las propiedades de seguridad son una preocupación principal para el desarrollo de sistemas de micropagos para evitar riesgos financieros para los comerciantes y también para garantizar la privacidad de los clientes. Por otro lado, la eficiencia y el costo de las transacciones individuales son factores críticos para el desarrollo de estos sistemas. Sin embargo, la eficiencia y la seguridad generalmente se oponen, por lo que los micropagos deben proporcionar una compensación entre estos requisitos.

En los últimos tiempos hemos visto el auge de las criptomonedas, unos medios digitales de intercambio, que utilizan la criptografía y están controladas por bases de datos descentralizadas e inalterables como son las cadenas de bloques o blockchain. Éstas presentan muchas ventajas respecto a las monedas fiduciarias, pero también presentan limitaciones, como son los elevados costes de las transacciones realizadas sobre las blockchain. Además, las operaciones realizadas en la blockchain no son completamente anónimas, ya que pueden ser rastreadas al ser visibles públicamente. Aún así, los usuarios no pueden ser fácilmente identificados, ya que utilizan direcciones que son seudónimos.

Podemos realizar transacciones instantáneas fuera de la blockchain utilizando los llamados canales de pago, que nos permitirán hacer transacciones instantáneas con criptomonedas. Éstos nos permiten superar los problemas de la blockchain en cuanto a costes elevados y de escalabilidad, en términos de transacciones por segundo y de espacio ocupado en la misma. Los

canales de pago se basan en la inclusión de muchas operaciones en la misma transacción de la blockchain principal. Por ello, son una excelente solución de segunda capa que eliminan la dependencia directa de la blockchain.

A. Contribución

En [1] propusimos un esquema de micropagos novedoso, eficiente y seguro para pagar artículos de bajo valor asegurando la privacidad de los clientes. En el presente artículo proponemos una versión mejorada del protocolo basada en el uso de blockchain. El anterior sistema descrito en [1] utilizaba monedas específicas para cada comerciante en cambio, con la introducción de la blockchain, el nuevo sistema utiliza criptomonedas en un intercambio equitativo entre clientes y comerciantes para pagar el bien o servicio deseado. También, nuestra nueva propuesta evita el uso de un banco que administre las cuentas de los clientes y comerciantes gracias al despliegue de *smart contracts* en la blockchain.

El sistema crea un canal de pago, evita el doble gasto y el gasto excesivo, protege contra la falsificación y, además, permite a los clientes solicitar un reembolso seguro de la cantidad no utilizada. El comerciante puede cobrar las cantidades pagadas incluso antes del cierre del canal y este puede redirigirse, permitiendo la transferabilidad de las criptomonedas, en lo que se denomina una solución *multihop*.

B. Organización

El trabajo está organizado de la siguiente forma. Primero, en la siguiente sección, describimos brevemente las características y los requisitos de seguridad de los micropagos, las criptomonedas, los canales de pago y los protocolos de compra, analizando el trabajo relacionado. En la sección III damos una visión general de la propuesta y los actores implicados. Luego, en la sección IV, definimos el protocolo de microcompra equitativa. En la sección V presentamos una descripción general de seguridad del protocolo. Finalmente, en la sección VI, el trabajo incluye las conclusiones y trabajos futuros.

II. SITUACIÓN ACTUAL

En [1] se presenta un esquema de micropagos eficiente y seguro, donde se describe un intercambio equitativo entre la moneda y el bien o servicio deseados, asegurando el anonimato y la imposibilidad de rastrear a los clientes. En el presente artículo se pretende mejorar dicho esquema, aprovechando las ventajas que nos ofrece actualmente la blockchain.

En [5], los autores introdujeron dos esquemas simples de micropago, PayWord y MicroMint, que utilizaban cadenas de hashes fuera de línea, pero no utilizaba aún la tecnología blockchain.

En [3] ya se propuso un sistema descentralizado que utilizaba la blockchain Bitcoin, mediante el cual las transacciones se envían a través de una red de canales de micropago. Estas transferencias de valor se realizan fuera de la blockchain, que las validaba posteriormente.

La irrupción de la blockchain Ethereum [6] supuso un paso más en las tecnologías de cadenas de bloques, al presentar el paradigma de una máquina transaccional singleton descentralizada con estado compartido. Esta tecnología ha propiciado la propuesta de multitud de soluciones con canales de micropago.

Por ejemplo, los canales de micropagos sobre la red Ethereum ya han sido tratados en [2], tratando de mejorar su rendimiento y coste. El protocolo propuesto en este artículo escala logarítmicamente con la capacidad del canal, utilizando una variante del árbol Merkle, y no requiere que el pagador bloquee todo el saldo en la creación del canal.

En [4] Di Ferrante también mostró cómo construir canales de pago en Ethereum usando sólo ”50 líneas de código”, entre el pagador y el beneficiario. El smart contract implementado permite verificar las firmas digitales de los pagos realizados fuera de la blockchain utilizando el código de operación $ecrecover$, que devuelve la dirección del firmante.

[7] también utiliza la tecnología Ethereum, proponiendo el intercambio de tokens híbrido descentralizado (HEX) para combinar los beneficios de los intercambios de tokens CEX (centralizados) y DEX (descentralizados). Este HEX amplía las soluciones existentes al agregar una nueva capa de canal de pago para beneficiar a los comerciantes frecuentes y aliviar la congestión de transacciones pendientes.

En [8] se propuso el uso de una red de canales de pago multihop y anónimo. La solución se basa en la criptografía de curva elíptica (ECC) y se ha demostrado que es segura al tiempo que logra la privacidad de la ruta de pago y el anonimato del remitente y el receptor.

También existen otras redes como Raiden [9], una red para transacciones instantáneas, que opera sobre la plataforma Ethereum y es un análogo a la idea de Lightning Network. Esta red pretende resolver el problema de escalabilidad que tiene actualmente la red Ethereum.

De todas formas, a pesar de que las redes de canales de pago hayan intentado mitigar los problemas de escalabilidad inherentes a las redes blockchain, éstas aún no brindan garantías significativas de seguridad y privacidad, como se demuestra en [11].

III. VISIÓN GENERAL DEL SISTEMA

El protocolo está formado por diferentes fases. En la primera de ellas se realiza la configuración del sistema y la generación de claves. A continuación, el usuario comprador selecciona el comercio en el que desea realizar las compras y solicita el servicio o producto a adquirir. En la siguiente fase se realiza la apertura del canal. Una vez el canal se halla abierto pueden realizarse las operaciones de compra, que incluyen el pago y el envío del producto o servicio adquirido. Finalmente el sistema cuenta con la fase de cobro o liquidación de canal o, alternativamente, la fase de transferencia del canal (multihop exchange), donde se realiza la reconversión del canal para la operación de compra con un nuevo usuario.

A. Actores

Los actores que intervienen en el protocolo de compra son el comprador C i el vendedor M. Además de éstos, el protocolo viene regulado por un smart contract (SC) desplegado en la blockchain que regula las operaciones sobre el canal de pago abierto entre C y M guardando los datos del intercambio en la blockchain. Al tratarse de un sistema de canales transferibles también pueden intervenir múltiples vendedores. Para definir la transferencia del canal se denominará N al segundo vendedor.

B. Notación

La tabla I muestra la notación utilizada en la descripción del protocolo.

Tabla I
NOTACIÓN.

Notación	Descripción
C	Comprador
M	Vendedor
N	Vendedor en un canal transferido
SC	Smart Contract
S_{id}	Identificador del servicio
Γ	Conjunto de elementos del canal
W_{LC}	Generador de la cadena
W_{0C}	Identificador de la cadena
W_{0M}	Identificador para la liquidación de la cadena
c	Número de μ-monedas de pago.
v	Valor de lsa μ-monedas
T_{Exp}	Fecha de caducidad
Δ_{TD}	Período de depósito
Δ_{TR}	Período de reintegro.
K_s	Clave de sesión
$E_{Ks}[M]$	Cifrado Simétrico del mensaje M con la clave K_s
$Channel_{Id}$	Identificador del canal
$H()$	Función de Hash
Q	Cantidad asociada al canal de pago

IV. PROTOCOLO

De acuerdo con el apartado anterior, describimos el protocolo de pago en sus diferentes fases: *Solicitud del servicio* por parte de un cliente, *Apertura* del canal de pago, *Compra* (realización del intercambio entre el pago por un producto o servicio), *Liquidación del canal* de pago y, finalmente, *Transferencia del canal* de pago para otros usos (*multihop*). Todas estas fases, a excepción de la fase de *Compra*, se realizan con comunicaciones *onchain*. Esto significa que las acciones, que realizan los distintos actores, las llevan a cabo a través de llamadas a funciones al smart contract desplegado para regular el canal de pago. Este smart contract controlará que cada actor sólo pueda realizar las acciones que le corresponden y dejará constancia de las mismas en la blockchain.

A continuación se describen cada una de las fases del protocolo propuesto:

A. Solicitud de Servicio

En esta fase el cliente debe seleccionar un servicio de les ofrecidos por los diferentes vendedores. Por tanto, cada vendedor M tiene publicado su lista de servicios disponibles, S_{Idi}. Luego, C selecciona un vendedor M y un servicio de la lista de servicios que ofrece. El servicio seleccionado se identifica mediante S_{id}. El cliente decide cuantas μ-monedas c quiere utilizar en el canal. Cada μ-moneda tendrá un valor v que se puede determinar en función del precio del servicio seleccionado (el precio de S_{id} será un múltiplo de v). El cliente envía estos datos a M que genera W_{LM} (donde $L = 2c + 1$) y aplica la función de hash L veces sobre este ítem $W_{0M} = H^L(W_{LM})$. El vendedor M transmite W_{0M} a C para que pueda abrir el canal de pago para este servicio. Cabe reseñar que esta cadena se genera para permitir liquidaciones parciales del canal. En caso de permitir una única liquidación entonces seria suficiente que M aplicase la función de hash una única vez sobre el generador para obtener el identificador.

B. Apertura del Canal

C procede a la apertura del canal para micropagos transfiriendo la cantidad $Q = c * v$ al smart contract, recordemos que c es el número de μ-monedas que quiere depositar en el canal y que v es su valor. Este realiza una comprobación del saldo de la transacción de C y almacena la cantidad Q.

Entonces C genera W_{LC}, donde $L = 2c + 1$ y aplica sobre él L veces una función de hash para obtener W_{0C}. Llamaremos $W_{(L-1)C}$ al resultado de aplicar la función de hash a W_{LC}. El último elemento de la cadena se denominará W_{0C}. Dentro de esta cadena los elementos con subíndice impar representarán μ-monedas mientras que los elementos con subíndice par representarán la prueba de pago de la anterior μ-moneda. Se ha decidido que la cadena de M tenga la misma longitud que la cadena de C para simplificar la notación aunque una cadena con la mitad de elementos seria suficiente ya que la cadena de M no contiene ítems de pago intercalados con ítems de prueba.

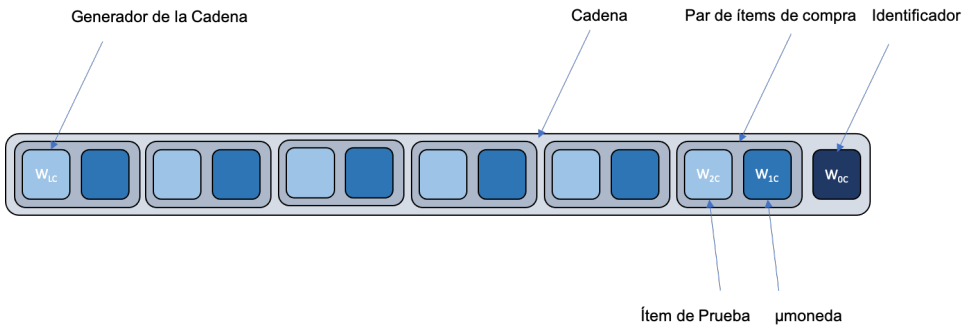


Fig. 1. Estructura de la cadena de ítems de compra

T_{Exp} es la fecha de caducidad del canal, Δ_{TD} define un período posterior a la fecha de caducidad en el cual puede realizarse la transferencia del contenido del canal a la cuenta del receptor del pago o bien la trasferencia del canal a un nuevo vendedor, mientras que Δ_{TR} representa un período para el reintegro del dinero restante, no utilizado para pagos en el canal, por parte de C y para una posible transferencia del canal por parte de M (ver Figura 2).

Con estos elementos, el comprador C genera el elemento $\Gamma = (W_{0M}, S_{Id}, 2c, v, T_{Exp}, \Delta_{TD}, \Delta_{TR}, W_{0C})$. Posteriormente, el comprador C llama a una función del smart contract para publicar en la blockchain el elemento Γ y, de esta manera, hacer efectiva la creación del canal de pago. Adicionalmente, el smart contract configura el contador $j = 0$ que simboliza el número de secretos revelados para el gasto/cobro de las μ-monedas. Identificaremos un canal en concreto por $Channel_{Id} = H(\Gamma)$.

C. Compra (Intercambio de pago y producto/servicio)

Pueden realizarse tantas operaciones de pago como $\mu-$monedas contenga el canal antes de T_{Exp}. Por ello M, a través del smart contract, comprobará antes de cada pago que la fecha actual es menor que la fecha de caducidad.

El proceso de compra está formado por tres pasos, formando un intercambio equitativo entre la $\mu-$moneda y el producto o servicio. En cada operación de compra pueden utilizarse una o varias $\mu-$monedas, de modo que la suma de sus valores sea igual a la cantidad a pagar en la operación de compra.

La operación de compra equitativa consta de tres pasos que se ejecutan off-chain, entre C y M: envío de las $\mu-$monedas, envío del producto o servicio y envío de la prueba de pago. Las $\mu-$monedas se revelarán en orden inverso a su creación en la cadena de hash. Para este proceso se utilizará una clave secreta de sesión K_s compartida por C y M.

1) Paso 1.

   C envía a M el mensaje $m_1 = [E_{Ks}[W_{iC}], Channel_{Id}]$. Al recibir el mensaje, M descifra W_{iC}, verifica la fecha y el identificador del canal, es decir comprueba que el canal esté abierto en el smart contract.

   A continuación, el vendedor M comprueba que no se ha producido reutilización de μ-moneda, comprobando que $i > j$, siendo i el número de orden de la μ-moneda utilizada en el pago y j el número de orden de la prueba de la última μ-moneda utilizada. Además, M verifica que W_{iC} pertenece a la cadena $H^{i-j}(W_{iC}) == W_{jC}$. Si se trata del primer pago realizado en este canal, la comprobación se hará del siguiente modo: $H^i(W_{iC}) == W_{0C}$. Una vez finalizadas estas verificaciones, M guarda S_{Id}, $Channel_{Id}$, W_{iC} y $(j = i)$.

2) Paso 2.

   M envía el producto o servicio mediante el mensaje m_2=$E_{Ks}[service/product]$. Al recibir el mensaje, C descifra m_2 y prepara la prueba adjunta a la μ-moneda.

3) Paso 3.

   C envía la prueba asociada a la μ-moneda de forma cifrada, junto con el índice correspondiente a su posición en la cadena. m_3 = $E_{Ks}[W_{(i+1)C}, (i+1)]$. El vendedor M descifra y verifica $W_{(i+1)C}$. Esta verificación se realiza aplicando una función de hash sobre la prueba y se verifica que el resultado se corresponde con la μ-moneda utilizada para el pago W_{iC}. Finalmente, M guarda el valor $W_{(i+1)C}$ y actualiza $j = i + 1$.

D. Liquidación del canal

Para que los fondos asociados con las operaciones de compra se transfieran a M, este debe realizar la operación

Fig. 2. Ciclo de vida del Canal de pago

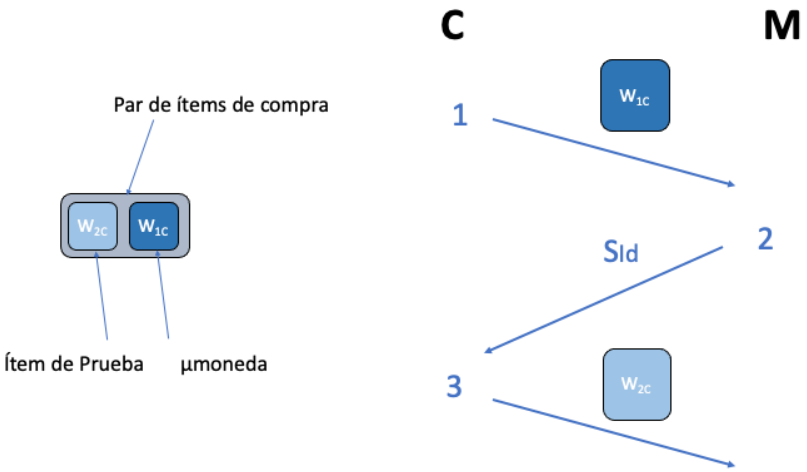


Fig. 3. Subprotocolo de Compra

de cobro. Esta operación puede realizarse una vez se hayan utilizado todas las μ-monedas asociadas al canal, pero también puede efectuarse si M ha recibido una o diversas μ-monedas asociadas con el canal de pago, aunque no se trate de su totalidad. En caso extremo, M podría realizar el cobro por cada μ-moneda individual gastada. Para esto M debe revelar cada uno de los W_{iM} en orden inverso a su creación. (Cabe destacar que la finalidad de los canales de pago es reducir costes y por tanto la eficiencia del sistema se consigue mediante el cobro simultáneo de diferentes $\mu-$monedas.)

Esta operación se realiza on-chain y para ello M accede al smart contract, el cual realizará la transferencia unificando la cantidad asociada a las μ-monedas reveladas. En un primer paso, el smart contract verificará que la fecha actual se encuentre en el período correcto. El período de cobro se encuentra entre T_{Exp} y $T_{Exp} + \Delta_{TD}$. Sin embargo, M también puede ejecutar la función de cobro antes de T_{Exp} para realizar un cobro parcial con las μ-monedas recibidas (en este escenario se podrán producir a posteriori más pagos con μ-monedas del canal). En la ejecución de la función de cobro del smart contract, M utilizará como parámetros los valores de $Channel_{Id}$, W_{kM}, W_{kC}, k, siendo k el índice de la última micromoneda recibida (o de la última moneda que se quiere cobrar).

El smart contract verifica que:

- $k > j$. Mediante esta comprobación se evita la reutilización de micromonedas.
- $W_{jM} == H^{k-j}(W_{kM})$. Esta comprobación permite verificar que el usuario que ejecuta la función es el destinatario del canal de pago.
- $W_{jC} == H^{k-j}(W_{kC})$. Se verifica que la μ-moneda pertenece a éste canal.

Una vez realizadas todas las comprobaciones, el smart contract realiza la transferencia del balance asociado a los micropagos a M. Para ello determina el valor en función de la cantidad de μ-monedas transferidas y del valor de cada una de ellas establecido en el canal. El número de μ-monedas transferidas se determina a partir del índice proporcionado por M. Si el índice es par se transferirá el valor de $(k-j)/2$ mientras que si es impar se transferirá el valor de $(k-j-1)/2$ ya que se tiene en cuenta que los índices incluyen las μ-monedas de prueba. Después de esta operación el smart contract tiene que actualizar el índice $j = k$.

E. Transferencia del Canal

El protocolo se ha diseñado teniendo en cuenta el interés en conseguir canales transferibles (o multihop), es decir, canales en los que el dinero pueda cambiar de manos en diferentes ocasiones sin necesidad de una transferencia de fondos hacia una wallet particular.

Con el fin de realizar una descripción más sencilla, pero sin quitar generalidad al funcionamiento del protocolo, supondremos que M va ha utilizar μ-monedas del mismo valor en ambos canales (el protocolo podría fácilmente

extenderse al uso de μ-monedas de valor diferente). El subprotocolo para que M pueda transferir los fondos del canal utilizado con C para realizar pagos a través de un nuevo canal a otro vendedor N es el siguiente:

- M solicita al smart contract que las μ-monedas recibidas pasen a otro canal, para efectuar pagos a N invocando a una función del smart contract que gestiona los canales de pago. Esta función solo se podrá ejecutar de forma alternativa a la función de liquidación de canal. Para ello N genera W_{0N} y M genera W'_{0M}, de forma análoga a como se generó W_{0M} y W_{0C} en el proceso de apertura del canal. $W_{0N} = H^n(W_{nN})$ y $W'_{0M} = H^n(W'_{nM})$. Como ya hemos mencionado, en este caso el valor de las μ-monedas v es el mismo que en el canal transferido. Por otra parte el numero de μ-monedas del canal puede ser igual o inferior al número de micromonedas del canal original $c' <= j$. Por tanto, el valor de n utilizando es $n = c' * v$.

- M genera Γ'= $(W_{0N}, S_{IdN}, 2c', v, T'_{Exp}, \Delta'_{TD}, \Delta'_{TR}, W'_{0M})$.

- M puede utilizar las μ-monedas del canal para realizar compras en N, siguiendo el subprotocolo de compra descrito anteriormente. Finalmente, si así lo desea, N recibiría la transferencia del dinero mediante el subprotocolo de liquidación del canal.

F. Reembolso

Una vez se ha liquidado el canal, las μ-monedas no utilizadas en operaciones de pago se retornarán a C. Se ha decidido que este reembolso no sea efectuado de forma automática por el Smart Contract para permitir a C decidir si quiere obtener el reembolso o bien reconvertir el canal para el pago a un nuevo vendedor.

El procedimiento de transferencia de canal puede ser utilizado también por C para crear un nuevo canal para cambiar las μ-monedas no gastadas y utilizarlas en un nuevo canal con otro vendedor M^*. Este cambio puede efectuarse en la ventana comprendida entre $(T_{Exp} + \Delta_{Td})$ y $(T_{Exp} + \Delta_{Td} + \Delta_{Tr})$.

Estas operaciones de transferencia del canal permiten reducir los costes asociados las operaciones de transferencia sobre la blockchain.

V. PROPIEDADES

En esta sección de describirán brevemente las principales propiedades del protocolo de gestión de canales para compras equitativas.

A. Anonimato

Los usuarios C y M, (y N si es el caso) acceden al smart contract mediante sus direcciones de blockchain.

Sin embargo, el canal se asocia al conocimiento de unos determinados valores. El usuario que conozca los elementos de la cadena de liquidación asociada al canal será el usuario que podrá efectuar la liquidación o la transferencia del canal. Por tanto, el usuario ejecuta el protocolo sin la necesidad de identificación.

El hecho de tratarse de canales multi-hop, con posibilidad de transferencia del canal sin que se produzca un movimiento de saldos sobre la blockchain permite que un usuario haya participado en las operaciones de compra sin la necesidad de que su cuenta en la blockchain haya emitido ni recibido ninguna transacción.

B. Equidad

La operación de compra se considera una aplicación de intercambio equitativo de valores. Por una parte, se realiza un pago y por otra se proporciona un servicio o producto.

En este protocolo el intercambio se realiza sin la intervención de ninguna tercera parte de confianza. Para realizar este intercambio equitativo se procede a la ejecución de un subprotocolo off-chain, donde los mensajes se intercambian directamente entre los usuarios C y M.

El procedimiento consta de tres pasos y podría interrumpirse sin llegar a completar la operación. Los posibles puntos de interrupción son los siguientes.

- Después del primer paso del intercambio, donde C proporciona la $\mu-$moneda W_{iC}, M no sigue con el protocolo y no proporciona el producto o servicio.

  - En este caso C no enviará el elemento de prueba $W_{(i+1)C}$.

  - La operación de compra se considera no realizada. C no recibe el producto o servicio y M no puede incluir la $\mu-$moneda en la operación de liquidación del canal o de transferencia ya que desconoce el elemento de prueba asociado a la última $\mu-$moneda.

  - En este primer escenario, efectivamente M no va a poder cobrar la última $\mu-$moneda pero sí que podría cobrar todas las anteriores. Por tanto, el cliente si quiere eliminar el riesgo de perder parcialmente el último $\mu-$pago, al no recibir el correspondiente servicio, puede ajustar el valor de la $\mu-$moneda a cada servicio y eliminar completamente este conflicto. No obstante, es

previsible que M no tenga este tipo de conductas ya que por cometer este $\mu-$fraude perdería las siguientes ventas al cliente y, por tanto, correría el riesgo de perder más que lo que ha ganado con este pequeño fraude. En todo caso, se trata de un riesgo controlado y regulado por el cliente, depende de él anular o no este potencial problema.

- Después del segundo paso del intercambio, C no sigue el protocolo y no proporciona el elemento de prueba asociado a la $\mu-$moneda, $W_{(i+1)C}$..
  - En este caso la compra se considera realizada.
  - C dispone del elemento adquirido mientras que M no puede incluir la $\mu-$moneda en las operaciones de liquidación o transferencia del canal.
  - Al realizar un nuevo pago con el canal, C utilizará la $\mu-$moneda $W_{(i+2)C}$. A partir de esta $\mu-$moneda, M puede derivar el elemento de prueba de la operación de compra anterior $W_{(i+1)C}$.
  - En este escenario un comprador únicamente puede dejar sin validación una única $\mu-$moneda por canal, lo cual entra dentro de los riesgos aceptables en operaciones de micropago. Por otra parte, el fraude asociado a la creación de canales para el robo de $\mu-$monedas individuales se descarta teniendo en cuenta que la creación del canal implica unos costes de ejecución que hacen inviable conseguir un beneficio de la creación de un canal para proceder al robo de una $\mu-$moneda.

C. Transferibilidad

El protocolo de compra se basa en el uso de un canal de pago establecido sobre blockchain. Este canal permite el pago desde un comprador C a aun vendedor V. En un escenario simple de ejecución M liquidará el canal después de recibir los pagos con las $\mu-$monedas asociadas al canal. Esta liquidación representará una transacción sobre la blockchain.

Sin embargo, con el objetivo de incrementar la eficiencia del sistema se ha diseñado el protocolo teniendo en cuenta la posibilidad de que los canales de pago sean *multihop*. En un canal *multihop* las $\mu-$monedas pueden cambiar de manos diversas veces sin necesidad de que cada cambio de manos represente una transacción en la blockchain.

La operación de transferencia del canal permite realizar los múltiples saltos del canal *multihop*. Si reas la recepción de las $\mu-$monedas el receptor, en lugar de realizar la liquidación del canal, opta por ejecutar la función de transferencia puede reorientar el canal para realizar pagos con el hacia un nuevo receptor. Los canales pueden prepararse para un número limitado de transferencias, ya que una transferencia ilimitada podría desembocar a una pérdida de eficiencia del sistema y debe tenerse en cuenta que en el escenario habitual de ejecución los canales no son transferidos más de un número reducido de veces.

El protocolo también contempla la posibilidad de que un usuario comprador no utilice todas las $\mu-$monedas del canal antes de la fecha de expiración del canal. Estos fondos pueden reembolsarse en la cuenta del comprador mediante una operación de transacción programada en el smart contract. Esta operación no se ha automatizado proporcionando al comprador la posibilidad de realizar una redirección del canal hacia un nuevo vendedor.

D. Imposibilidad de Falsificación, Sobregasto y Doble Gasto

El protocolo se ha diseñado para impedir la falsificación de las $\mu-$monedas así como su reutilización o el uso de más $\mu-$monedas de las establecidas en el canal.

- Todas las $\mu-$monedas recibidas como medio de pago en el protocolo de compra deben pertenecer al canal asociado. Esta comprobación se realiza mediante la operación $H^i(W_{iC}) == W_{0C}$, donde W_{0C} debe coincidir con el valor almacenado dentro del elemento Γ del canal.
- La reutilización de una $\mu-$moneda se evita mediante la comprobación $H^{i-j}(W_{iC}) == W_{jC}$, donde i es el índice de la $\mu-$moneda actual y j el índice de la $\mu-$moneda utilizada en el último pago en el canal, por lo que i debe ser mayor que j.
- La sobreutilización se evita con dos procedimientos. Por una parte, el dinero correspondiente a las $\mu-$monedas asociadas al canal queda depositado en el smart contract antes de iniciarse las operaciones de compra. Por otra parte el número de $\mu-$monedas asociadas al canal es limitado y la última $\mu-$moneda del canal viene determinada por el elemento $W_{(L-1)C}$. El vendedor no aceptará ninguna $\mu-$moneda que requiera una validación en la cadena de hash con más de L operaciones.

E. Coste Reducido

Un coste por pago reducido es una característica fundamental de los $\mu-$pagos ya que nunca debe superar el beneficio asociado a ese pago y mucho menos al valor de la cantidad transferida.

El protocolo presentado en este trabajo reduce considerablemente el cose de una operación de pago

mediante criptomonedas mediante la creación de un canal de pago. Aunque el número de operaciones de pago que pueden realizarse es de c operaciones por canal, el número de transferencias sobre la blockchain se reduce a uno canal, si no se contemplan las operaciones de transferencia.

Cuanto mas a largo término sean las relaciones entre comprador y vendedor más largas podrán ser las cadenas de $\mu-$monedas asociadas al canal y mayor será la eficiencia del sistema.

Por otra parte, el diseño de canales multihop permite reducir el número de transferencias reales sobre la blockchain, ya que un canal transferible no requiere liquidación después de el uso entre los dos primeros usuarios.

VI. CONCLUSIONES

En este trabajo hemos presentado un esquema de microcompras, es decir un sistema de intercambio equitativo entre un micropago y el acceso al producto o servicio comprado. Los micropagos son pagos de cantidades muy reducidas donde el margen de beneficio es escaso con lo que el sistema no debe tener unos costes asociados que reduzcan sustancialmente o eliminen totalmente ese margen de beneficio. Para ello los mecanismos de seguridad y privacidad asociados a los micropagos deben diseñarse en consecuencia.

En el caso de las criptomonedas la solución para implementar micropagos se encuentra en diseñar protocolos de capa dos [12] en los que se evite que cada operación de pago represente una transacción en la blockchain. Un canal, una vez abierto, permite la realización de operaciones off-chain que no se materializaran sobre la blockchain hasta el momento de liquidación del canal.

El artículo presenta un protocolo de gestión de canales mediante smart contracts. El sistema se basa en el protocolo de micropagos sin blockchain presentado en [1]. Igual que en el protocolo de [1] el pagador construye una cadena de elementos donde se intercalan micromonedas y elementos de prueba que permitirán otorgar atomicidad a las operaciones de compra.

El protocolo está formado por una fase inicial de solicitud de servicio seguido por una operación on-chain en la que se crea el canal. El canal no está asociado a un vendedor concreto, sino que este dispondrá de un ítem que le autentifica como el liquidador veraz del canal. A continuación, la operación de compra se realiza tantas veces como sea necesario de forma off-chain. Finalmente, el vendedor puede liquidar el canal con lo que se procederá a la transacción de transferencia de fondos sobre la blockchain.

Hemos añadido al protocolo básico las operaciones de transferencia del canal y reembolso para incrementar la eficiencia del sistema y evitar transacciones innecesarias.

En el artículo se realiza un análisis informal de las propiedades del sistema, incluyendo anonimato, equidad, transferibilidad, imposibilidad de falsificación, sobregasto y reutilización de monedas así como una valoración de la eficiencia del sistema.

Como trabajo futuro se pretende estudiar la reversibilidad del canal y realizar las modificaciones necesarias al protocolo para permitir pagos bidireccionales sobre el canal. Por otra parte, se realizará la implementación del protocolo sobre una blockchain de tipo ethereum mediante la programación de los smart contracts que se encargaran de las operaciones on-chain del protocolo, como la creación y la liquidación del canal.

AGRADECIMIENTOS

El proyecto FeltiCHAIN (RTI2018-097763-B-I00) está financiado por: FEDER/Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, (MCI/AEI/FEDER, UE).

REFERENCIAS

[1] Isern-Deyà, A. P., Payeras-Capellà, M. M., Mut-Puigserver, M., Ferrer-Gomila, J. L. (2013). "Anonymous and Fair Micropayment Scheme with Protection against Coupon Theft." International Journal of Adaptive, Resilient and Autonomic Systems (IJARAS), 4(2), 54-71. doi:10.4018/jaras.2013040103

[2] Galal H.S., ElSheikh M., Youssef A.M. (2019) "An Efficient Micropayment Channel on Ethereum". In: Pérez-Solà C., Navarro-Arribas G., Biryukov A., Garcia-Alfaro J. (eds) Data Privacy Management, Cryptocurrencies and Blockchain Technology. DPM 2019, CBT 2019. Lecture Notes in Computer Science, vol 11737. Springer, Cham. https://doi.org/10.1007/978-3-030-31500-9_13

[3] Poon, Joseph, and Thaddeus Dryja. "The bitcoin lightning network: Scalable off-chain instant payments." (2016).

[4] Di Ferrante, M. "Ethereum payment channel in 50 lines of code." Medium, June (2017).

[5] Rivest, Ronald L., and Adi Shamir. "PayWord and MicroMint: Two simple micropayment schemes." International workshop on security protocols. Springer, Berlin, Heidelberg, 1996.

[6] Wood, Gavin. "Ethereum: A secure decentralised generalised transaction ledger." Ethereum project yellow paper 151.2014 (2014): 1-32.

[7] Luo, Xuan, et al. "A payment channel based hybrid decentralized ethereum token exchange." 2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). IEEE, 2019.

[8] Tripathy, Somanath, and Susil Kumar Mohanty. "Mappcn: Multihop anonymous and privacy-preserving payment channel network." International Conference on Financial Cryptography and Data Security. Springer, Cham, 2020.

[9] https://raiden.network/

[10] https://blog.althea.net/universal-payment-channels/

[11] G. Malavolta, P. Moreno-Sanchez, C. Schneidewind, A. Kate, and M. Maffei, "Anonymous multi-hop locks for blockchain scalability and interoperability" in Network and Distributed System Security Symposium (NDSS), 2019.

[12] Lewis Gudgeon, Pedro Moreno-Sanchez, Stefanie Roos, Patrick Mccorry, and Arthur Gervais. "SoK: Layer-Two Blockchain Protocols". In International Conference on Financial Cryptography and Data Security, pages 201-226. Springer, 2020.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Generación automática de firmas para detección de ciberataques basados en URI

R. Estepa Alonso*, J. Diaz-Verdejo†, A. Estepa Alonso*, G. Madinabeitia*, F. J. Muñoz*
* Dpt. Ingeniería Telemática, Escuela Superior de Ingenieros, Univ. de Sevilla
C/ Camino de los Descubrimientos s/n, 41092 Sevilla (Spain)
E-mail: {rafa,aestepa,german,javi }@trajano.us.es
† Dpt. Teoría de Señal, Telemática y Comunicaciones, CITIC, Univ. de Granada
C/ Periodista Daniel Saucedo Aranda, s/n, 18071 Granada (Spain)
E-mail: jedv@ugr.es

La mayor parte de los sistemas de detección de intrusiones (IDS) operativos se basan en el uso de firmas que permiten identificar ataques conocidos. La dependencia de estos IDS con la actualización de las bases de datos de firmas constituye una de sus mayores limitaciones, siendo de interés el desarrollo de sistemas que posibiliten la generación automática o supervisada de firmas.

En el presente trabajo se evalúa experimentalmente un sistema para la generación de firmas a partir de un IDS basado en anomalías propuesto en un trabajo previo. También se desarrolla y evalúa un sistema automatizado para la selección del punto de operación óptimo del generador de firmas. Los resultados preliminares de este trabajo en curso muestran que se pueden generar firmas nuevas que aumenten la capacidad de detección del IDS basados en firmas o patrones conocidos (SIDS) controlando el número de falsos positivos introducidos.

***Palabras Clave*—Cybersecurity, Intrusion Detection, Automatic signatures generation, Web-based attacks**

I. INTRODUCCIÓN

La necesidad de proteger los equipos y redes de ciberamenzas es cada vez más notoria y relevante. Uno de los elementos clave en la seguridad de los sistemas y redes son los denominados sistemas de detección de intrusiones (IDS, del inglés *Intrusion Detection Systems*) [1], que emiten alertas a partir de la observación de los diversos eventos que ocurren en la red o los sistemas a proteger. Los IDS generan alertas según dos modos de operación básicos: basado en *firmas* (SIDS, del inglés *Signature-based IDS*), que identifican un patrón malicioso preestablecido denominado firma, como por ejemplo una secuencia dentro de la URI de una petición HTTP; o basados en anomalías (AIDS, del inglés *Anomaly-based IDS*), que identificación de comportamientos anómalos, dando lugar a los IDS.

Los SIDS son sistemas muy extendidos en la actualidad, dado que permiten detectar ataques ya conocidos con una fiabilidad y coste computacional razonables. Como es lógico, el adecuado comportamiento de los SIDS depende fuertemente de la disponibilidad y calidad de las firmas, que deben ser generadas y actualizadas periódicamente. Por tanto, estos sistemas resultan inadecuados para detectar ataques novedosos, o de día cero (*0-day*), por no existir firmas para los mismos. Sin embargo, éstos representan un porcentaje importante del total de ataques y, sobre todo, generan un fuerte impacto. La solución pasaría por la generación de las firmas correspondientes, pero este problema es recursivo, ya que para poder generar la firma es necesario detectar previamente el ataque, por lo que debe utilizarse algún procedimiento alternativo. De ahí el interés de desarrollar sistemas que sean capaces de generar las firmas de forma automática o semiautomática.

Como hemos mencionado anteriormente, los AIDS [1] constituyen una aproximación diferente a la detección de ataques y son potencialmente capaces de detectar ataques *0-day*. Su rendimiento dependerá de su capacidad de aprender y discriminar el comportamiento normal/anómalo. En entornos IT, donde en ocasiones no hay un patrón claro de comportamiento del usuario, esta tarea se adivina compleja, lo que propicia la aparición de numerosos falsos positivos (FP), siendo ésta una de las mayores limitaciones de los AIDS en la actualidad.

Son múltiples los trabajos en los que se ha propuesto el uso de AIDS para identificar ataques y, a partir de ellos, generar las firmas correspondientes para los SIDS [2]. Para ello, se necesita no sólo determinar si se está desarrollando un ataque, sino también identificar los elementos significativos del mismo, que serán los asociados a la firma. El interés de esta aproximación reside en la mayor facilidad de uso e implementación de los SIDS, y en la posible capacidad de generalización de las firmas

así obtenidas, eliminando o reduciendo significativamente la intervención de los expertos. Su utilidad, no obstante, vendría limitada por las tasas de FP a las que podrían dar lugar estas nuevas firmas.

En un trabajo previo [3] se ha propuesto un sistema automático para la generación de firmas en el contexto de ataques basados en URI (véase Sección II). El AIDS subyacente se basa en [4], que modela las URI en base a una aproximación markoviana que permite identificar los elementos asociados en mayor medida a la clasificación como ataque y, consecuentemente, proponer firmas para los mismos. Los resultados obtenidos evidencian la posibilidad de conseguir una generación de firmas adecuada, pero son fuertemente dependientes del punto de operación del sistema, que es ajustado de forma manual en un procedimiento que puede resultar complejo.

En el presente trabajo en curso pretendemos explorar las capacidades de dicha propuesta en un escenario operativo real que incluye varios servidores que cooperan para establecer las nuevas firmas. Para ello se abordan propuestas y mejoras en tres aspectos relevantes. En primer lugar, se plantea un sistema automático de selección del punto de operación óptimo para la generación de las firmas, analizando el impacto de los FP sobre las reglas generadas y, consecuentemente, sobre el uso de las mismas en el escenario real. Por otra parte, se plantean diversas técnicas para la selección y agrupación de las firmas a partir de los segmentos identificados como asociados a ataques. Finalmente, se analizará la capacidad de generalización de las firmas a partir de su distribución a otros servicios diferentes a aquel en el que se ha inferido. El objetivo final es el desarrollo de un sistema global de generación y distribución de firmas para ataques basados en URI. Este trabajo se está llevando a cabo en el ámbito de un proyecto de colaboración con una empresa andaluza del sector de SmarCities, que proporcionará datos reales obtenidos durante operación.

El presente artículo se estructura como sigue. En primer lugar, en el Apartado II se presentará brevemente la técnica SSM y el trabajo previo en el que se basa la presente propuesta. El Apartado III describe la arquitectura general del sistema propuesto y aborda el problema del ajuste automático del punto de operación, presentándose el escenario utilizado para estas pruebas y los resultados experimentales obtenidos en el Apartado IV. Finalmente, en el Apartado V se presentan las conclusiones y se esbozan los desarrollos y resultados preliminares relativos a la agrupación de firmas y su distribución.

II. Generación de firmas

A continuación, describiremos brevemente los fundamentos de la técnica utilizada y su aplicación a la generación de firmas de ataques [3].

A. Detección de anomalías en URI

La técnica utiliza un autómata de estados finitos probabilístico para representar las instancias de un protocolo con estructura sintáctica en sus cargas útiles (en nuestro caso las URI de HTTP) mediante su segmentación en *palabras*. De acuerdo al estándar RFC 3986, un URI, U_k, debe presentar una estructura sintáctica de la forma:

"http://"host[":"port][abs_path["?"query]]

siendo posible su segmentación, a partir de los delimitadores estándar, en un conjunto de L palabras, $w_k = \left\{w_1^k, w_2^k, \cdots, w_L^k\right\}$, asociadas a cada uno de los campos (en nuestro caso sólo son de interés los campos *abs_path* y *query*, formada por los pares *atributo, valor*).

A partir de un conjunto de URI, es posible establecer un *diccionario*, $D = \{(w_i, f_i)\}$, compuesto por todas las palabras observadas, w_i y su frecuencia relativa de observación, f_i. De esta forma, dado un URI de entrada U_k compuesto por una secuencia de palabras, w^k y un diccionario previamente estimado, es posible asignar un *índice de anomalía*, $A_s(U_k)$, a partir de la probabilidad estimada para cada una de dichas palabras [5]:

$$A_s(U_k) = -log\left(\frac{1}{L}\sum_{i=1}^{L}\log(f_i^k)\right) \tag{1}$$

Este índice será positivo y tanto mayor cuanto menor sea la probabilidad de la secuencia observada. De esta forma, se podrá clasificar un URI como normal o anómalo de acuerdo al *umbral de detección*, θ, como

$$Clase(U) = \begin{cases} Normal & si\ A_s(U) < \theta \\ Anomalo & si\ A_s(U) \geq \theta \end{cases} \tag{2}$$

Por otra parte, esta aproximación plantea un problema de *entrenamiento insuficiente* relacionado con la posible aparición de palabras que no han sido observadas durante el proceso de entrenamiento y que, en consecuencia, tendrían asociada una probabilidad nula. Para solucionarlo se establece una probabilidad fija mínima para cualquier palabra observada, denominada *probabilidad de fuera de vocabulario*, p_{OOV}.

B. Generación de firmas

El modelado anteriormente descrito permite evaluar la probabilidad de normalidad de las distintas palabras que componen la URI, por lo que, dada una URI que se determina anómala (ataque), es posible identificar y seleccionar los segmentos que contribuyen en mayor proporción a dicha clasificación. De esta forma, se delimitan y extraen las palabras o secuencias de palabras que superan el denominado *umbral de generación de firma para un segmento*, ϕ, incluyendo los delimitadores correspondientes. Cada uno de estos fragmentos será candidato a formar parte de una nueva firma.

Por otra parte, el propio índice de anomalía de una URI es indicativo del grado de normalidad de la misma, por lo que, para minimizar el posible impacto de los FP, se establece un *umbral de generación de firmas*, Ψ, de tal forma que únicamente las URI cuyo índice de anomalía supere dicho umbral serán consideradas en el proceso de generación de firmas. En consecuencia, dado un URI, U, se determina que un segmento t es anómalo y se incorpora a una firma si se cumple

$$(A_s^t(U) \geq \phi)\ \wedge\ (A_s(U) \geq \Psi),\ con\ \Psi > \theta \tag{3}$$

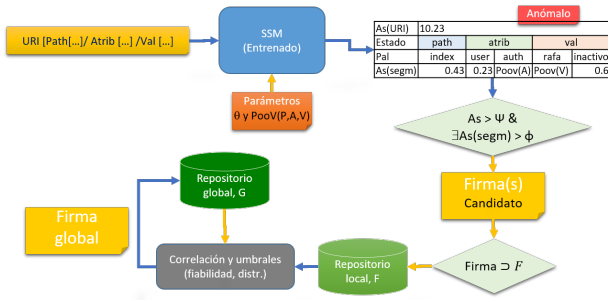


Fig. 1. Funcionamiento del sistema propuesto

siendo $A_s^t(U)$ el índice de anomalía del segmento.

La operación del sistema propuesto en este trabajo se esquematiza en la Fig. 1. Por un lado, cada uno de los AIDS desplegados y entrenados con su tráfico local evalúan las URI de entrada y, para aquellas suficientemente anómalas, extraen los segmentos candidatos a firmas, que serán agrupados convenientemente en una nueva firma integrada en un repositorio local de firmas. Como se puede observar, a partir de los modelos entrenados y ajustados en varios servidores se infieren repositorios de firmas locales que son agrupadas y analizadas para extraer un repositorio global con firmas válidas para todos los servidores. La generación de un repositorio global cooperativo de firmas será abordado en las siguientes fases del proyecto en curso, centrándose este trabajo en el sistema generador de firmas.

III. Ajuste de umbrales de la generación de firmas

Para la extracción de las firmas locales es necesario ajustar experimentalmente el sistema para seleccionar el punto óptimo de operación, que influirá en la tasas finales de detección y de falsos positivos. Consecuentemente, es necesario ajustar 3 parámetros: θ, ϕ y Ψ, ya que el valor de p_{OOV} depende del conjunto de entrenamiento. Así, el valor del umbral de generación de firma para un segmento, ϕ, debe ser inferior al de la probabilidad mínima registrada en el diccionario, esto es, $\phi < min(\{f_i\})$, para asegurar que las palabras que constituyen la firma no han sido observadas previamente. Así mismo, parece lógico pensar que las URI candidatas a generación de firmas sean un subconjunto de aquellas detectadas como anómalas, lo que exige que se cumpla $\theta < \Psi$. También resulta coherente que, para controlar el número de FP que pueden dar lugar a firmas, haya que ajustar el valor de Ψ. A continuación, proponemos un procedimiento de ajuste del umbral de generación de firmas en el que acotamos la tasa máxima de FP aceptada. Este algoritmo parte de la suposición de que la tasa de FP objetivo que tengamos en el conjunto de entrenamiento será similar a la que obtendremos durante la explotación del sistema.

A. Ajuste automático del valor de Ψ

El objetivo del mecanismo de ajuste que se propone en este trabajo es explorar un espacio de búsqueda de valores para Ψ a fin de que la tasa de FP conseguida con las firmas

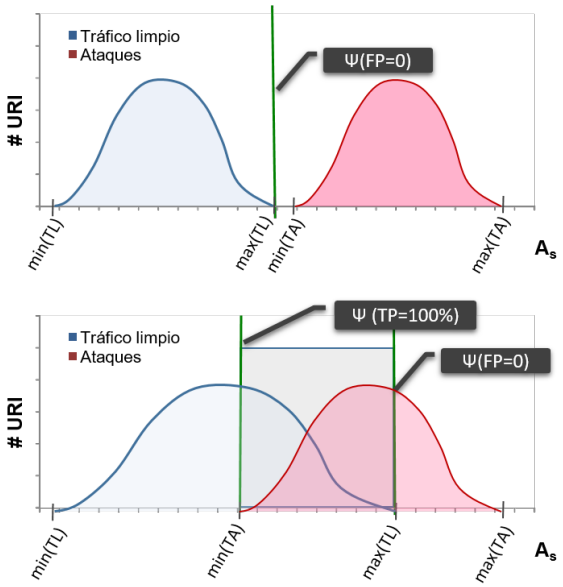


Fig. 2. Casos para el histograma de A_s.

no sobrepase un umbral determinado por el operador del servicio. En primer lugar, podemos determinar cotas para el valor de Ψ, umbral de generación de firma, a la vista de los índices de anomalía registrados durante la fase de entrenamiento.

Dado un dataset de entrenamiento con tráfico limpio (TL) y otro con tráfico de ataques (TA), es de esperar que el histograma de los índices de anomalía responda a una de las dos situaciones mostradas en la Fig. 2. En el primer caso (parte superior), que correspondería a la situación ideal, el tráfico limpio y el de ataque presentan una gran diferencia en sus diccionarios, resultando que $max(A_s(TL)) < min(A_s(TA))$, lo que implica que si elegimos $\Psi > max(A_s(TL))$ no tendremos ningún FP en el entrenamiento y detectaremos todos los ataques. Desafortunadamente, el segundo caso es el más habitual e implica que $max(A_s(TL)) > min(A_s(TA))$, por lo que valores de Ψ en el rango $[min(A_s(TA)), max(A_s(TL))]$ generarán una tasa de falsos positivos en el entrenamiento.

Así pues, el ajuste de Ψ se realizará durante el entrenamiento, evaluando iterativamente la tasa de FP encontrada en el TL cuando se utilizan las firmas generadas[1] para valores crecientes de Ψ. Esto se puede hacer con un algoritmo que parte de un valor inicial $\Psi = min(A_s(TA))$, que generará la tasa de FP máxima posible, que se computará a partir de TL. Si dicha tasa es menor que la tasa de FP objetivo, el algoritmo se detendrá, en otro caso, se incrementará el valor de Ψ y se volverá a evaluar en una nueva iteración. El resultado final será el valor de Ψ que cumple que la tasa de FP que introducen las nuevas firmas es menor que el valor objetivo.

IV. Resultados experimentales preliminares

A continuación, se presentan los resultados experimentales obtenidos relativos a la capacidad de detección y

[1]A tal efecto se ha desarrollado una sencilla herramienta SIDS denominada *InspectorLog*, que permite aplicar las firmas generadas a las URI.

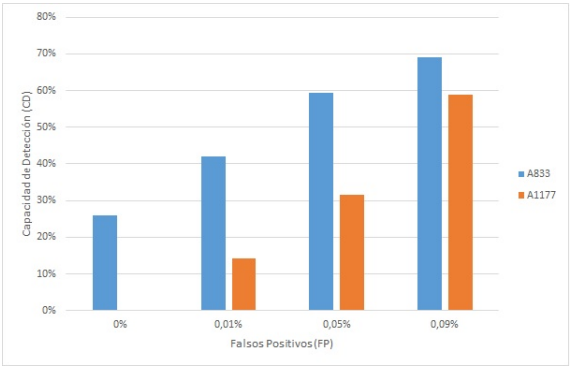


Fig. 3. Capacidad de detección de las firmas en diversos puntos de operación del AIDS.

el ajuste de umbrales. El valor de ϕ se ha ajustado a $0,9 \cdot min(\{f_i\})$, cumpliendo así la restricción de que un segmento anómalo no puede haber sido visto en el tráfico limpio. Para la experimentación se ha utilizado:

- Trafico limpio (TL): proveniente de 1 semana de tráfico real del servicio ProxyWeb de una empresa, que denominaremos H, que cuenta con 289 505 peticiones GET. Se han realizado 4 particiones para entrenamiento, test y validación.
- Tráfico de ataques (TA): se han utilizado dos dataset con 833 y 1 177 URI de ataques, respectivamente, generadas a partir de las vulnerabilidades encontradas en la base de datos CVE (*Common Vulnerabilities and Exposures*) aplicables a servidores HTTP del año 2018 [5].

El primer experimento realizado utiliza el algoritmo de ajuste de Ψ propuesto anteriormente para obtener firmas con distintos umbrales de FP tolerados en el AIDS: 0%, 0,01%, 0,05% y 0,09%. Para ello se entrena el sistema con una de las cuatro particiones y se evalúa con el resto, promediando los resultados según un esquema *leave-one-out*. Los resultados finales obtenidos para las firmas generadas con los distintos dataset de ataques se muestran en la Fig. 3.

En esta figura se puede observar que, a mayor FP objetivo mayor capacidad de detección de las firmas generadas. Con respecto a los FP detectados, siempre fueron inferiores al FP objetivo del algoritmo, tomando los valores de 0%, 0.001%, 0.007%, 0.023% para los FP objetivos 0%, 0.01%, 0.05% y 0.09% respectivamente. Estos resultados avalan la hipótesis de que la tasa de FP generados por el AIDS será siempre superior a la de las firmas obtenidas.

El siguiente experimento realizado consistió en explorar los límites del sistema cuando se establece la tasa de FP a 0, para observar la capacidad máxima de detección obtenida. En la Tabla I se pueden observar los resultados para el dataset de 833 ataques. Vemos que entrenando con el tráfico limpio H1 (primera partición) tan sólo somos capaces de detectar un 33,73% de los ataques, que generarían 33 firmas. Las distintas particiones de TL empleadas (H1-H4) dan lugar a diferentes valores. Para cada experimento se muestra el valor óptimo de Ψ determinado por el algoritmo, el máximo valor que podría tomar (columna rango), la capacidad de detección de ataques, los falsos positivos encontrados y el número de firmas generadas.

Tabla I
RESULTADOS DE GENERACIÓN DE FIRMAS CON DIFERENTES PARTICIONES.

Exp	Ψ	rango	CD(%)	FP(%)	N. Firmas
H1.833	16.31	17.36	33,73	0	33
H2.833	16.27	17.31	33,73	0	66
H3.833	16.83	17.33	2	0	20
H4.833	16.29	17.33	33,7	0	33

V. CONCLUSIONES

La generación automatizada permite mejorar la capacidad de detección de los SIDS. En este artículo se ha evaluado el rendimiento de un sistema generador de firmas en el contexto de ataques en la URI así como un método para el ajuste de umbrales y reducción de FP. También se han presentado algunos resultados preliminares dentro de los límites de espacio asociados al tipo de trabajo (en curso). Los resultados muestran la capacidad de detección de ataques novedosos que no eran detectados mediante las firmas disponibles sin incrementar la tasa de FP del SIDS. Actualmente estamos trabajando con datasets de mayor tamaño que permiten seguir desarrollando y mejorando el sistema, así como en el uso cruzado de las firmas para estudiar la capacidad de generalización.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto 2020/00000172 dentro del programa de Proyectos singulares de actuaciones singulares de transferencia en los CEI en las áreas RIS3 de la Junta de Andalucía.

REFERENCIAS

[1] N. Moustafa, J. Hu, J. Slay, "A holistic review of Network Anomaly Detection Systems: A comprehensive survey", Journal of Network and Computer Applications,(128)33?55, 2019.
[2] S. Kaur, M. Singh, "Automatic attack signature generation systems: A review", IEEE Secur. Priv., (11)54–61, 2013.
[3] P. Garcia-Teodoro, J.E. Diaz-Verdejo, J. Tapiador, R. Salazar-Hernandez, "Automatic generation of HTTP intrusion signatures by selective identification of anomalies", Computers and Security, (55)159–174, 2015.
[4] J. M. Estévez-Tapiador, P. García-Teodoro, J. E. Díaz-Verdejo, "Detection of web-based attacks through Markovian protocol parsing", Proc. IEEE Symp. on Computers and Communications, 2005.
[5] R. Estepa, J.E. Díaz-Verdejo, A. Estepa, G. Madinabeitia, "How Much Training Data Is Enough? A Case Study for HTTP Anomaly-Based Intrusion Detection", IEEE Access, 8:44410–44425, 2020.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Detección de ataques de red mediante clasificación de flujos empleando L-momentos

Jesús Galeano-Brajones*, Jose J. Rico-Palomo*, Mihaela I. Chidean†, Javier Carmona-Murillo*

* Departamento de Ingeniería de Sistemas Informáticos y Telemáticos. Universidad de Extremadura. 06006.
† Departamento de Teoría de la Señal y Comunicaciones. Universidad Rey Juan Carlos. 28942.

jgaleanobra@unex.es, jjricopal@unex.es, mihaela.chidean@urjc.es, jcarmur@unex.es

El incremento continuo de dispositivos conectados a Internet en los últimos años, junto con el aumento de las aplicaciones y servicios, han propiciado que la tarea de clasificación del tráfico de red sea esencial en cualquier entorno, tanto para cuestiones de gestión de red como de seguridad. Con la llegada de las redes 5G han aparecido nuevos retos relacionados con la seguridad debido a este gran volumen de tráfico y la diversidad de servicios disponibles para los usuarios. Además, también han surgido nuevas tecnologías basadas en software y virtualización que permiten un control más dinámico de la red. En este contexto, este artículo propone una metodología novedosa para procesar los flujos de red mediante el cálculo de los L-momentos estándar y su posterior clasificación para la detección de anomalías y amenazas en la red. Además, se ha desarrollado un *testbed* con el que poder experimentar con cualquier conjunto de datos y estos estadísticos. Los resultados obtenidos tras la experimentación con este *testbed* muestran que los L-momentos estándar resultan especialmente útiles para procesar los flujos de red en tiempo real, consiguiendo que los algoritmos de clasificación obtengan unos resultados de calidad muy elevados.

***Palabras Clave*—detección de ataques, L-momentos, ciberseguridad, clasificación, machine learning**

I. Introducción

En los últimos años tanto los dispositivos conectados a Internet como el volumen de tráfico de red han crecido de manera significativa. La clasificación de flujos de red se ha convertido en una tarea fundamental con el avance de tecnologías como la 5G y el aumento de dispositivos IoT (*Internet of Things*). Según *Cisco Annual Internet Report 2020* [1], el 66% de la población global tendrá acceso a Internet en 2023, es decir, 5.3 mil millones de usuarios, lo que implica un crecimiento enorme de datos moviéndose por la red. Además, el número de brechas de seguridad y el total de registros expuestos por brecha continúa creciendo, pasando de 7.9 millones de ataques de denegación de servicio distribuidos (DDoS, *Distributed Denial of Service*) a los 15.4 millones previstos para 2023.

5G es la quinta generación de redes móviles. Es un nuevo tipo de red diseñada para conectar prácticamente a todos y a todo, incluyendo máquinas, objetos y dispositivos. 5G está destinado a ofrecer velocidades de datos de hasta varios gigabits por segundo, latencias ultrabajas, una mayor fiabilidad, una capacidad de red masiva, mayor disponibilidad y una experiencia de usuario más uniforme para más usuarios. Un mayor rendimiento y una mayor eficiencia potencian nuevas experiencias de usuario y conectan nuevas industrias [2].

Con 5G se prevén más casos de uso que en las generaciones de redes móviles anteriores, pudiendo clasificarse en tres tipos de servicios: *Enhanced Mobile Broadband* (eMBB), *Ultra-Reliable and Low Latency Communications* (URLLC) y *Massive Machine-Type Communications* (mMTC). El primero busca ofrecer a los usuarios enlaces con tasas de datos más uniformes, mayor capacidad y menor latencia, como por ejemplo en Realidad Virtual y Realidad Aumentada. El segundo está más relacionado con la Industria 4.0, ya que ofrece enlaces ultra-fiables y de baja latencia, como el control remoto de infraestructuras críticas o vehículos, así como el apoyo a procedimientos médicos a distancia. En el último, 5G pretende conectar un número masivo de dispositivos, como sensores embebidos y dispositivos IoT, gracias a la capacidad de adaptar la tasa de datos de los dispositivos para mejorar la eficiencia del ancho de banda, la eficiencia energética y para múltiples escenarios de movilidad, proporcionando soluciones de conectividad de bajo coste. Sin embargo, la creciente popularidad de estos dispositivos los convierte en el objetivo de los atacantes. Por ello es necesario garantizar la seguridad de estos dispositivos y redes desarrollando mecanismos más avanzados capaces de detectar amenazas de seguridad y mitigarlas. Este es un

reto importante para los dispositivos IoT, ya que manejan información sensible y muchos no suelen implementar medidas de seguridad adecuadas [3], lo que los hace perfectos para integrarlos en *botnets* que realicen ataques DDoS a servicios de mayor envergadura.

El desarrollo del 5G se basa en varias tecnologías habilitadoras clave, como las redes definidas por software (SDN, *Software-Defined Networking*) y la virtualización de funciones de red (NFV, *Network Function Virtualization*). SDN es un paradigma de red propuesto para acabar con las limitaciones de las infraestructuras de red actuales, rompiendo la integración vertical mediante la separación de la lógica de control de la red (plano de control) de los *routers* y *switches* subyacentes que reenvían el tráfico (plano de datos) e incrementa la flexibilidad de la red. NFV es una tecnología que desacopla las funciones de red del hardware subyacente, permitiendo el reemplazo de hardware dedicado, propietario y caro con dispositivos de red basados en software. NFV también permite que instancias de funciones virtuales se compartan entre varios clientes [4]. Ambas tecnologías permiten el desarrollo de nuevos mecanismos de seguridad y el despliegue de estas soluciones en redes de este tipo, como por ejemplo, el despliegue de algoritmos de clasificación de flujos de red en un controlador SDN.

Este trabajo propone una metodología novedosa para la clasificación de flujos de red, específicamente para la detección de anomalías en la red en base al procesamiento de flujos mediante los L-momentos estándar y posterior clasificación mediante algoritmos de *Machine Learning* (ML). Hasta donde conocemos, por el momento no se ha explorado en la literatura este método para procesar los flujos. También se ha implementado esta metodología mediante el desarrollo de un *testbed* con el que se puede experimentar con cualquier *dataset*. Este *testbed* permite calcular los L-momentos y L-momentos estándar para cada una de las características de los flujos del *dataset* y aplicar diferentes algoritmos de clasificación con el fin de obtener resultados que indiquen la calidad de las clasificaciones. Además, este artículo es la evolución natural de los siguientes trabajos: en primer lugar, en el artículo [5] se caracterizan las amenazas del conjunto de datos UNSW-NB15 [6] mediante los diagramas de L-momentos estándar; en segundo lugar, en el artículo [7] se realiza la detección y mitigación de ataques DoS (*Denial of Service*) y DDoS utilizando la entropía de características extraídas de los flujos de red. Además, se ha utilizado una *stateful-SDN* para gestionar la red y el conjunto de datos Bot-IoT [8].

El resto del artículo se organiza como sigue. En la Sección II se proporciona una revisión de la literatura relacionada con este trabajo. Los antecedentes teóricos y tecnológicos base para la metodología propuesta se describen en la Sección III. La Sección IV detalla la evaluación experimental realizada. Finalmente, la Sección V incluye las principales conclusiones alcanzadas, así como las líneas de investigación futura que quedan abiertas.

II. Trabajo relacionado

En esta sección se proporciona una revisión de la literatura relacionada con este artículo. Por esta razón, se ha dividido en dos subsecciones. La primera describe las principales líneas de investigación relacionadas con la aplicación de la teoría de los L-momentos en diferentes campos de conocimiento. La segunda muestra algunos de los trabajos más relevantes relacionados con la clasificación de flujos de red, específicamente de clasificación de anomalías de red.

A. Aplicación de los L-momentos

La teoría de los L-momentos ha sido ampliamente utilizada por la comunidad científica desde su propuesta en 1990 [9], aunque no ha sido explorada en el campo del tráfico de red y la telecomunicación. Recientemente, los L-momentos han tenido diferentes campos de aplicación, como climatología, aplicaciones de radar y bioingeniería.

Los L-momentos han sido utilizados principalmente en la regionalización del clima y para el cálculo del SPI (*Standarized Precipitation Index*) y SPEI (*Standarized Precipitation Evapotranspiration Index*) [10], dos índices climáticos muy utilizados para el estudio de sequías. También han sido utilizados para estimar parámetros de distribuciones de probabilidad que modelan la máxima velocidad del viento en escala temporal [11].

En el contexto de la teoría de redes complejas, los L-momentos de una distribución específica se han utilizado para crear un test multivariante de hipótesis para detectar la sutil degradación de nodos o aristas [12].

En los campos de ingeniería o bioingeniería, los L-momentos se han utilizado para clasificar objetivos en aplicaciones de radar [13] y también para clasificar glóbulos blancos [14].

Por último, solo encontramos dos trabajos relacionados con la clasificación de tráfico. Por un lado, en [15], los *trimmed L-moments* [16] se utilizan para estimar la distribución generalizada de Pareto, la cual es utilizada para modelar distribuciones de colas pesadas, como el modelado de tráfico de red [17]; por el otro, la teoría de los L-momentos se utiliza para caracterizar flujos de red entre legítimos y anómalos [5]. Este último puede ser considerado un trabajo preliminar al trabajo llevado a cabo en este artículo.

B. Clasificación de flujos de red

La clasificación de flujos de red se ha convertido en una tarea fundamental para la monitorización de flujos, especialmente para la detección de anomalías y flujos pertenecientes a amenazas. Esta clasificación ha sido ampliamente estudiada desde las siguientes cuatro perspectivas: basada en puertos, inspección profunda de paquetes (DPI, *Deep Packet Inspection*), basados en el *payload* y enfoques estadísticos (ver Fig. 1).

La técnica basada en puertos es una técnica sencilla y rápida que utiliza la asociación de los puertos de la cabecera TCP/UDP con puertos bien conocidos asignados

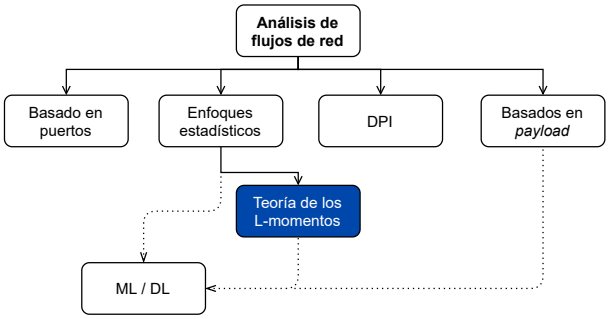


Fig. 1. Taxonomía de la clasificación de flujos de red

por la IANA (*Internet Assigned Numbers Authority*) [18]. Pero la amplia proliferación de nuevos servicios y aplicaciones que utilizan puertos no asignados por la IANA ha incrementado significativamente, como los dispositivos IoT que utilizan direcciones IP privadas o dinámicas y números de puertos variables [19], haciendo que esta técnica deje de ser útil.

DPI emergió como una alternativa a la técnica basada en puertos e incluso hoy en día es una de las más aceptadas y utilizadas. Esta técnica busca encontrar patrones o palabras clave en los paquetes de datos. Su mayor limitación es que solo es aplicable a paquetes que no están encriptados, además de los problemas para la privacidad de los usuarios [20]. Como alternativa, muchos investigadores proponen el uso de técnicas de ML puesto que evita muchos de estos inconvenientes [21].

La técnica de clasificación de flujos basada en el *payload* solo utiliza la información de la capa de aplicación. Esto solventa la dependencia de IP y puertos de la técnica basada en puertos. Normalmente se despliega junto a DPI [19], [22].

Junto con los enfoques estadísticos y las técnicas basadas en el *payload*, en los últimos años ha habido un aumento de las aplicaciones de Deep Learning (DL), incluso relacionadas con la clasificación de tráfico o flujos de red. Para ello, se han estudiado principalmente *Stacked Auto-Encoders* (SAE) [23], redes neuronales recurrentes (RNN, *Recurrent Neural Networks*) [24] y Redes Neuronales Convolucionales (CNN, *Convolutional Neural Netowkrs*) [25].

Las técnicas que emplean un enfoque estadístico utilizan parámetros independientes del *payload* como la duración de los flujos, el tiempo entre llegadas, la longitud de la cabecera de los paquetes, etc. Estos parámetros pueden utilizarse para alimentar diferentes modelos estadísticos, de ML o de DL. La metodología propuesta en este artículo utiliza este enfoque, analizando estadísticamente diferentes parámetros de los flujos mediante la teoría de los L-momentos y clasificándolos utilizando diferentes algoritmos de ML.

III. Metodología

En esta sección se presenta la metodología propuesta para la clasificación de flujos, comenzando con la definición de los L-momentos, incluidos los L-momentos estándar, y finalizando con la descripción del *testbed* desarrollado y el conjunto de datos utilizado en la experimentación.

A. L-momentos

Al igual que otros momentos estadísticos, los L-momentos caracterizan la geometría de distribuciones y resumen muestras. Son directamente análogos, es decir, tienen interpretación similar, a los momentos centrales (llamados *product moments* o *C-moments* en la literatura [26]). Los L-momentos son combinaciones lineales de diferencias de esperanzas de estadísticos de orden. Esta es la principal diferencia con los momentos centrales, los cuales están basados en potencias de diferencias con la media. Algunos beneficios de los L-momentos frente a los momentos centrales son:

- Los L-momentos son más robustos ante la presencia de datos atípicos, es decir, sufren menos por los efectos de la variabilidad de las muestras.
- Necesitan menos datos para estimar con errores bajos, lo que los hace muy útiles para estimaciones en tiempo real y para trabajar con L-momentos de orden alto.
- En contraste a los momentos centrales, los L-momentos son insesgados, es decir, no dependen del tamaño de la muestra.
- Están más cerca de su distribución normal asintótica en muestras finitas.

A continuación, se definen los L-momentos teóricos y muestrales, finalizando con el diagrama de L-momentos estándar.

1) L-momentos teóricos: Los estadísticos de orden de una variable aleatoria X para una muestra de tamaño n están formados por el orden ascendente $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$. Los L-momentos teóricos quedan definidos por:

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} E[X_{r-k:r}], \quad \forall\ r \geq 1 \tag{1}$$

donde r es el orden del L-momento y $E[X_{r-k:r}]$ es la esperanza del estadístico de orden $r-k$ de una muestra de tamaño r. Los primeros L-momentos teóricos pueden definirse en función de las esperanzas de los estadísticos de orden:

$$\begin{aligned}
\lambda_1 &= E[X_{1:1}] \\
\lambda_2 &= \frac{1}{2}E[X_{2:2}] - E[X_{1:2}] \\
\lambda_3 &= \frac{1}{3}E[X_{3:3}] - 2E[X_{2:3}] + E[X_{1:3}] \\
\lambda_4 &= \frac{1}{4}E[X_{4:4}] - 3E[X_{3:4}] + 3E[X_{2:4}] - E[X_{1:4}]
\end{aligned} \tag{2}$$

λ_1 se denomina *L-location* y λ_2 es *L-scale*, una medida de la variabilidad de la distribución. Además, las distribuciones útiles tienen una variabilidad distinta de

cero, por lo que $\lambda_2 > 0$. Con λ_2 se definen los momentos estándar teóricos:

$$\tau_r = \lambda_r/\lambda_2, \quad \forall\ r \geq 3 \tag{3}$$

Algunos momentos estándar tienen un nombre definido:

$$\begin{aligned} \tau_2 &= \lambda_2/\lambda_1 = \text{coeficiente de } \textit{L-variation} \\ \tau_3 &= \lambda_3/\lambda_2 = \textit{L-skewness} \\ \tau_4 &= \lambda_4/\lambda_2 = \textit{L-kurtosis} \end{aligned} \tag{4}$$

τ_2 es útil para variables aleatorias positivas ($X \geq 0$) y está acotado en $0 < \tau_2 < 1$. Para $r \geq 3$, todos los momentos estándar están acotados en $-1 < \tau_r < 1$. Y específicamente, *L-kurtosis* está acotada en $\frac{1}{4}\left(5\tau_3^2 - 1\right) \leq \tau_4 < 1$. Estas limitaciones hacen que los momentos estándar sean muy útiles puesto que permiten comparar distribuciones cuyos rangos son diferentes sin la necesidad de normalizar o reescalar.

2) L-momentos muestrales: Los L-momentos muestrales se calculan para una muestra de estadísticos de orden $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$. La muestra de estadísticos de orden se estiman simplemente ordenando los datos en orden ascendente. Los L-momentos muestrales se definen como:

$$\hat{\lambda}_r = \tfrac{1}{r}\tbinom{n}{r}^{-1} \sum_{i=1}^{n} \left[\sum_{j=0}^{r-1} (-1)^j \tbinom{r-1}{j} \tbinom{i-1}{r-1-j} \tbinom{n-i}{j} \right] x_{i:n}, \quad \forall\ r \geq 1 \tag{5}$$

Y los L-momentos estándar muestrales son:

$$\hat{\tau}_r = \hat{\lambda}_r/\hat{\lambda}_2, \quad \forall\ r \geq 3 \tag{6}$$

$$\begin{aligned} \hat{\tau}_2 &= \hat{\lambda}_2/\hat{\lambda}_1 = \text{coeficiente de } \textit{L-variation} \text{ muestral} \\ \hat{\tau}_3 &= \hat{\lambda}_3/\hat{\lambda}_2 = \textit{L-skewness} \text{ muestral} \\ \hat{\tau}_4 &= \hat{\lambda}_4/\hat{\lambda}_2 = \textit{L-kurtosis} \text{ muestral} \end{aligned} \tag{7}$$

3) Diagrama de L-momentos estándar: Como ya se ha mencionado, gracias a que los L-momentos estándar están acotados, se pueden comparar directamente diferentes distribuciones. Un diagrama de L-momentos estándar (*L-moments ratio diagram*, a partir de ahora LmomRD por sus siglas en inglés) es un gráfico que muestra los L-momentos estándar en cada dimensión. Lo más común es utilizar τ_3 frente a τ_4 debido a que estos son los L-momentos estándar que tienen un mayor soporte teórico, pero los LmomRD no están limitados a estos puesto que existen L-momentos estándar de mayor orden cuya definición es mucho más compleja. La Fig. 2 muestra el LmomRD por defecto con la representación de algunas distribuciones comunes. Las distribuciones con tres parámetros se representan como una línea, mientras que las distribuciones con dos parámetros se representan con un punto. Estas representaciones ayudan a interpretar los datos e identificar a qué distribución puede ajustarse una muestra real. Esta herramienta visual es utilizada en este trabajo y, para facilitar comparaciones entre diferentes resultados, las representaciones de las distribuciones más comunes estarán en todos los LmomRD.

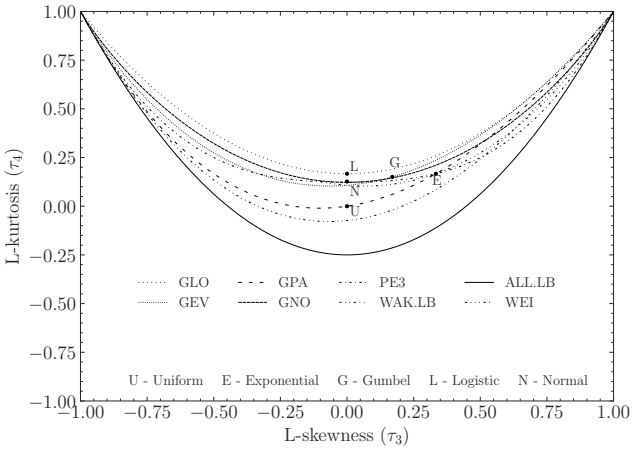


Fig. 2. Diagrama de L-momentos estándar de algunas distribuciones comunes (GLO: Generalized Logistic; GEV: Generalized Extreme Value; GPA: Generalized Pareto; GNO: Generalized Normal; PE3: Pearson Type 3 o Gamma; WEI: Weibull; WAK.LB: límite inferior de la distribución Wakeby; ALL.LB: límite inferior de cualquier distribución) [9]

B. Testbed

Con el objetivo de poder experimentar con los L-momentos y cualquier conjunto de datos disponible en la literatura, se ha desarrollado un *tesbed* que permite realizar esta experimentación de manera automática. La metodología seguida se muestra en la Fig. 3.

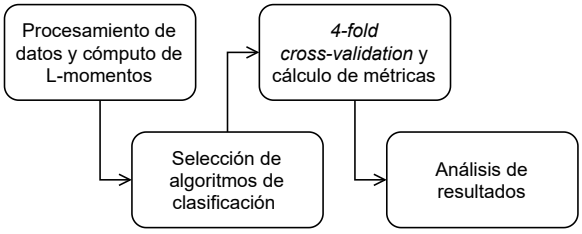


Fig. 3. Fases de la metodología

En primer lugar, se procesan los conjuntos de datos y se calculan los L-momentos y L-momentos estándar. En segundo lugar, se seleccionan aquellos modelos de ML con los que se quiere experimentar. En tercer lugar, se realiza el entrenamiento con *4-fold cross-validation* y se obtienen las métricas de calidad de las clasificaciones para, por último, realizar un análisis de los resultados. Desde un punto de vista más práctico, en la Fig. 4 se muestra el diagrama de flujo que define el comportamiento del *testbed*, donde los bloques de condición se corresponden con *flags* que guían el flujo en función de las necesidades del usuario:

- `compute_lmom`. Cómputo de los L-momentos y L-momentos estándar. Si se han calculado previamente, se cargan de los ficheros sin necesidad de volver a realizar el cómputo.

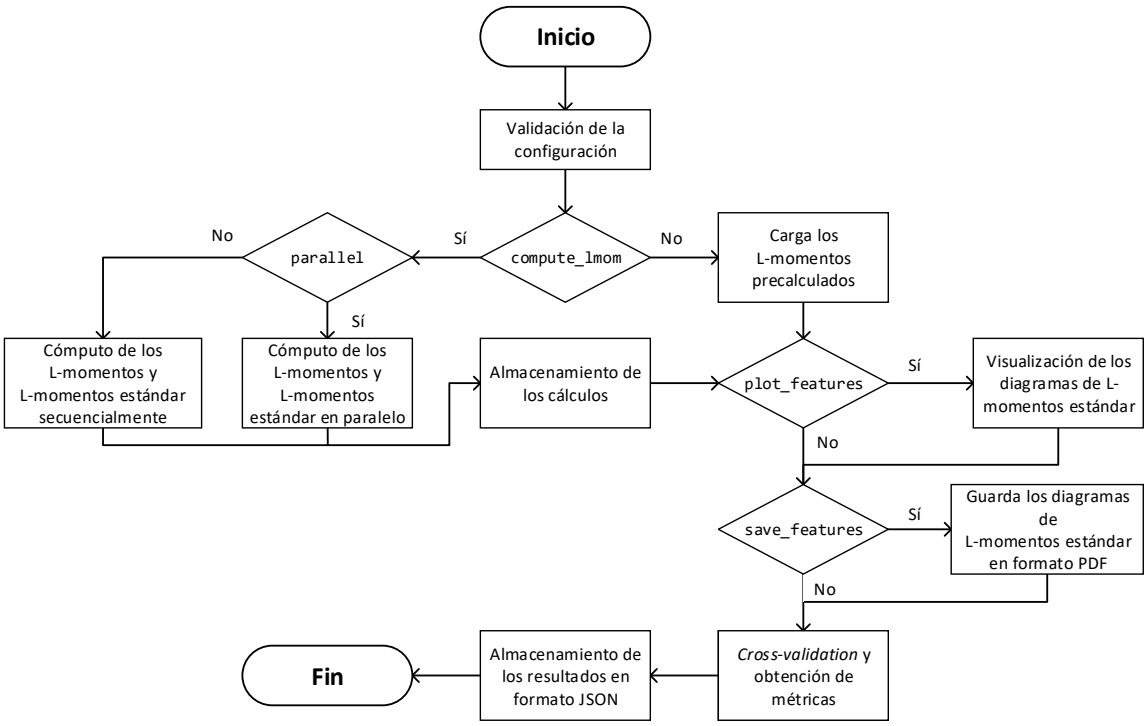


Fig. 4. Diagrama de flujo del *testbed*

- `parallel`. Cómputo de manera secuencial si *false* o en paralelo aprovechando todos los núcleos de la CPU si *true*.
- `plot_features`. Visualización de los LmomRD en tiempo real.
- `save_features`. Almacenamiento de los LmomRD si *true*.

C. CSE-CIC-IDS2018

Este conjunto de datos es un proyecto colaborativo entre el *Communications Security Establishment* (CSE) y el *Canadian Institute for Cybersecurity* (CIC) [27]. Se trata de un conjunto de datos generado de manera experimental que incluye siete escenarios de ataque: fuerza bruta, *Heartbleed*, *botnet*, DoS, DDoS, ataques web e infiltración en la red. Además, los creadores del conjunto de datos ofrecen distintos tipos de ficheros para su utilización: ficheros PCAP con todo el tráfico capturado, ficheros CSV etiquetados con 80 características de los flujos y *logs* del sistema de cada una de las máquinas que intervienen en el escenario.

IV. Resultados

En esta sección se describen algunas pruebas realizadas con el *testbed* descrito en la Sección III.B. En primer lugar, se muestra el efecto del valor n, es decir, el tamaño de la muestra, de manera visual mediante la utilización de los LmomRD. En segundo lugar, se expone el comportamiento del algoritmo kNN para clasificar los L-momentos estándar para distintos valores de n. Todas las pruebas se han realizado con el conjunto de datos CSE-CIC-IDS2018, específicamente con la captura del día 20/02/2018, la cual contiene flujos legítimos y flujos etiquetados como ataques DDoS LOIC HTTP. LOIC (*Low Orbit Ion Cannon*) es una aplicación desarrollada en C# para realizar ataques DoS y DDoS utilizando TCP, UDP y HTTP. Fue desarrollada durante el Proyecto Chanology, una serie de protestas en Internet promovidas por el grupo Anonymous contra la Iglesia de la Cienciología [28]. En este conjunto de datos, los ataques DDoS LOIC se realizan utilizando HTTP. Además, tras realizar un análisis de las características de los flujos que se encuentran en el conjunto de datos, hemos concluido que el intervalo de llegada entre paquetes es una buena característica para mostrar los resultados.

A. Efecto de n en los LmomRD

El tamaño de la muestra n es un parámetro muy importante a la hora de estimar los L-momentos y los L-momentos estándar. Si n es un valor bajo se necesitan menos valores en la muestra para calcular cada L-momento estándar, es decir, cada punto del LmomRD. Pero, aunque se necesitan menos valores para formar la muestra, los *clusters* generados por los L-momentos estándar de los flujos quedan menos concentrados, es decir, las etiquetas pueden quedar mezcladas. Desde el punto de vista de la monitorización o muestreo de los flujos de la red, un n bajo implica mayor precisión temporal para una misma frecuencia de muestreo y mayor rapidez en la detección de amenazas si los *clusters* no están demasiado dispersos, un punto clave para la seguridad de las redes. Pero a su vez implica una mayor imprecisión si n no tiene

un valor lo suficientemente elevado como para definir los *clusters*, es decir, se obtiene un número de falsos positivos y falsos negativos elevado. Esto se observa claramente en el LmomRD de la Fig. 5. En este caso se ha utilizado $n = 50$ y, para esta característica del flujo en concreto, es un valor excesivamente bajo puesto que visualmente es difícil diferenciar con buena precisión los flujos legítimos de los flujos de ataque.

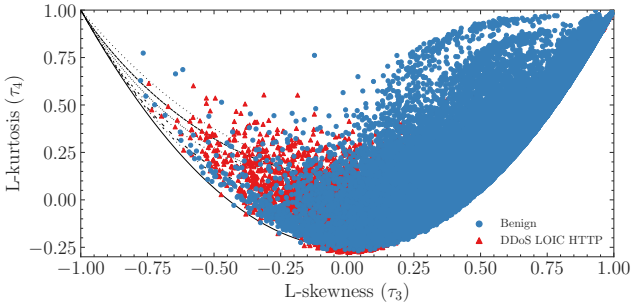


Fig. 5. Diagrama de L-momentos estándar para $n = 50$

Un aumento del tamaño de la muestra, como $n = 200$, implica una concentración más definida de los *clusters* (ver Fig. 6), consiguiendo diferenciarlos visualmente para cada etiqueta de los flujos aunque se observan algunos L-momentos estándar que serán falsos positivos y falsos negativos.

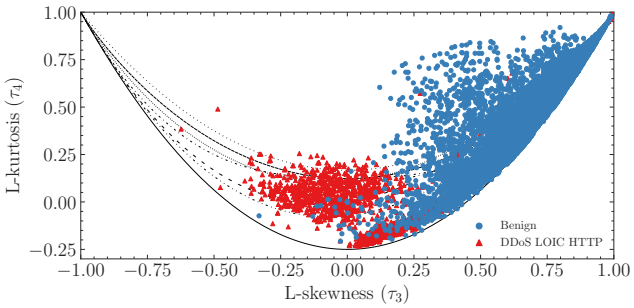


Fig. 6. Diagrama de L-momentos estándar para $n = 200$

Por último, con un valor mucho mayor, $n = 3200$, los *clusters* quedan mucho más definidos y concentrados (ver Fig. 7). Esto no tiene por qué ser así para cualquier característica de los flujos. Existen características que permiten diferenciar los flujos claramente y otras para las que los L-momentos estándar no son útiles. Es por ello que cada tipo de amenaza y característica de los flujos requiere un estudio exhaustivo, en primer lugar para conocer si la característica puede clasificarse y, en segundo lugar, para obtener un valor de n que satisfaga las necesidades de tiempos de muestreo y precisión de la detección.

B. Efecto de n en la clasificación

Como se ha comentado, el valor de n influye directamente en la precisión de la estimación de los L-momentos estándar. Por lo tanto, mientras que un n pequeño implica mayor dispersión, un n elevado conlleva una mejor concentración en *clusters* definidos,

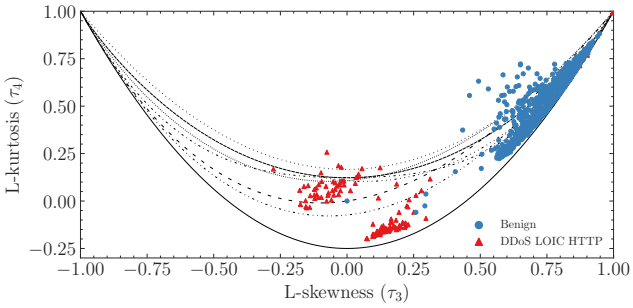


Fig. 7. Diagrama de L-momentos estándar para $n = 3200$

aumentando así la calidad de cualquier análisis posterior. Tras esta explicación, a continuación se muestra el efecto de este parámetro en la posterior clasificación de los L-momentos estándar con el algoritmo kNN y dos tipos de funciones de coste: uniforme (la distancia entre todos los puntos se pondera igual) y distancia (pesos definidos por el inverso de la distancia entre puntos). Además, el valor de k queda definido por $\sqrt{N}/2$, donde N es el número total de datos de entrenamiento.

En la Fig. 8 se muestra el resultado de la métrica *balanced accuracy* para diferentes valores de n. La elección de esta progresión geométrica permite representar resultados para un rango amplio de n a la vez que evita el comportamiento "ruidoso" de los mismos (p.e. se espera un resultado similar para $n = 800$ y $n = 850$ e incluir ambos en la figura dificultaría la visualización) La métrica *accuracy* es ampliamente utilizada en los problemas de clasificación, pero dado que el conjunto de datos está claramente desbalanceado, se utiliza *balanced accuracy* como métrica de calidad. Ambas métricas son equivalentes y *balanced accuracy* es más adecuado para conjuntos de datos como el de este trabajo.

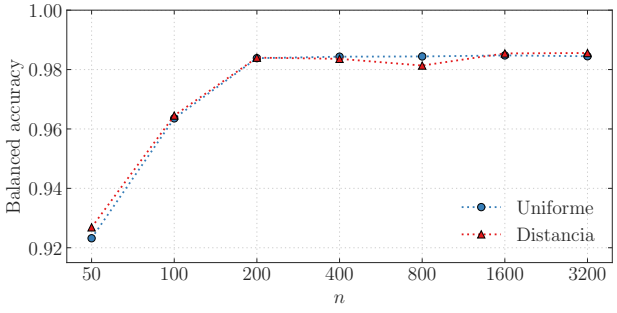


Fig. 8. *Balanced accuracy* para distintos valores de n

Para la característica considerada se puede observar que la clasificación obtiene unos resultados muy buenos incluso desde $n = 50$. De hecho, el resultado para este valor de n puede parecer no concordar con los visto en su LmomRD, sin embargo un análisis más detallado de los datos y los resultados obtenidos en la clasificación permite comprobar que la mayoría de los L-momentos estándar legítimos se encuentran en la zona de $\tau_3 > 0$ y el número de falsos negativos y falsos positivos de la matriz de confusión es mínimo en comparación con la cantidad

de L-momentos estándar presentes en el LmomRD.

V. Conclusiones

Debido al aumento del volumen de tráfico en la red y a la necesidad de clasificar los flujos, en este artículo se propone una novedosa metodología para realizar esta clasificación con el objetivo de detectar anomalías en la red y amenazas de seguridad. Esta metodología está basada en el uso de los L-momentos estándar, una herramienta que ha demostrado ser muy útil para esta tarea y que, hasta donde sabemos, aún no ha sido explorada en la literatura para propósitos similares a los nuestros. Tras la experimentación realizada, se puede concluir que los L-momentos son una herramienta muy apropiada para procesar los flujos de red y posteriormente entrenar algoritmos de clasificación con el fin de detectar ataques puesto que se obtiene una calidad de clasificación muy elevada. Además, esta metodología no solo diferencia entre flujos legítimos y maliciosos, sino que también diferencia entre tipos de ataques.

Finalmente, este trabajo tiene una amplia investigación futura. En primer lugar, los resultados de los LmomRD muestran formas de los *clusters* que deben ser comprendidas y justificadas puesto que pueden ayudar a la comunidad a comprender mejor el comportamiento estadístico de los ataques. En segundo lugar, en este trabajo nos hemos limitado a investigar con τ_3 y τ_4, pero como ya hemos mencionado, existen L-momentos estándar de mayor orden que también pueden ser de utilidad en este tipo de problemas. En tercer lugar, es necesaria una experimentación futura sobre entornos prácticos, como el despliegue de la metodología en un controlador SDN. Por último, los algoritmos de clasificación pueden ser estudiados con el objetivo de optimizar sus parámetros y así obtener la mayor calidad posible de los resultados.

Agradecimientos

Este trabajo ha sido financiado parcialmente por el Ministerio de Ciencia, Innovación y Universidades con el proyecto RTI2018-102002-A-I00, y por la Consejería de Economía e Infraestructuras de la Junta de Extremadura con el proyecto IB18003 y la ayuda GR18141.

Referencias

[1] U. Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.

[2] D. Lake, N. Wang, R. Tafazolli, and L. Samuel, "Softwarization of 5G Networks – Implications to Open Platforms and Standardizations," *IEEE Access*, pp. 1–1, 2021.

[3] F. Meneghello, M. Calore, D. Zucchetto, M. Polese, and A. Zanella, "IoT: Internet of threats? A survey of practical security vulnerabilities in real IoT devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8182–8201, 2019.

[4] Y. Li and M. Chen, "Software-Defined Network Function Virtualization: A survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[5] M. I. Chidean, J. Carmona-Murillo, R. H. Jacobsen, and Q. Zhang, "Network Traffic Characterization Using L-moment Ratio Diagrams," in *2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, pp. 555–560, IEEE, 2019.

[6] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *2015 military communications and information systems conference (MilCIS)*, pp. 1–6, IEEE, 2015.

[7] J. Galeano-Brajones, J. Carmona-Murillo, J. F. Valenzuela-Valdés, and F. Luna-Valero, "Detection and Mitigation of DoS and DDoS Attacks in IoT-Based Stateful SDN: An Experimental Approach," *Sensors*, vol. 20, no. 3, p. 816, 2020.

[8] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," *Future Generation Computer Systems*, vol. 100, pp. 779–796, 2019.

[9] J. R. Hosking, "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 1, pp. 105–124, 1990.

[10] J. Hosking and J. Wallis, "Some statistics useful in regional frequency analysis," *Water resources research*, vol. 29, no. 2, pp. 271–281, 1993.

[11] M. Fawad, T. Yan, L. Chen, K. Huang, and V. P. Singh, "Multiparameter probability distributions for at-site frequency analysis of annual maximum wind speed with L-moments for parameter estimation," *Energy*, vol. 181, pp. 724–737, 2019.

[12] F. Mohd-Zaid, C. M. Schubert Kabban, and R. F. Deckro, "A test on the L-moments of the degree distribution of a Barabási–Albert network for detecting nodal and edge degradation," *Journal of Complex Networks*, vol. 6, no. 1, pp. 24–53, 2018.

[13] R. Ginoulhac, F. Barbaresco, J.-Y. Schneider, J.-M. Pannier, and S. Savary, "Target Classification Based On Kinematic Data From AIS/ADS-B, Using Statistical Features Extraction and Boosting," in *2019 20th International Radar Symposium (IRS)*, pp. 1–10, IEEE, 2019.

[14] K. Al-Dulaimi, K. Nguyen, J. Banks, V. Chandran, and I. Tomeo-Reyes, "Classification of White Blood Cells Using L-Moments Invariant Features of Nuclei Shape," in *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, IEEE, 2018.

[15] J. Hosking, "Some theory and practical uses of trimmed L-moments," *Journal of Statistical Planning and Inference*, vol. 137, no. 9, pp. 3024–3039, 2007.

[16] E. A. Elamir and A. H. Seheult, "Trimmed L-moments," *Computational Statistics & Data Analysis*, vol. 43, no. 3, pp. 299–314, 2003.

[17] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," *A practical guide to heavy tails: statistical techniques and applications*, vol. 23, pp. 27–53, 1998.

[18] IANA, "Service Name and Transport Protocol Port Number Registry." https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml. [Online, accessed 3 June].

[19] H.-K. Lim, J.-B. Kim, K. Kim, Y.-G. Hong, and Y.-H. Han, "Payload-based traffic classification using multi-layer LSTM in Software Defined Networks," *Applied Sciences*, vol. 9, no. 12, p. 2550, 2019.

[20] G. Li, M. Dong, K. Ota, J. Wu, J. Li, and T. Ye, "Deep Packet Inspection Based Application-Aware Traffic Control for Software Defined Networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2016.

[21] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin, and J. Aguilar, "Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1988–2014, 2018.

[22] J. Zhang, Y. Xiang, Y. Wang, W. Zhou, Y. Xiang, and Y. Guan, "Network traffic classification using correlation information," *IEEE Transactions on Parallel and Distributed systems*, vol. 24, no. 1, pp. 104–117, 2012.

[23] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Computing*, vol. 24, no. 3, pp. 1999–2012, 2020.

[24] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, "BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset," *IEEE Access*, vol. 8, pp. 29575–29585, 2020.

[25] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "DISTILLER: Encrypted traffic classification via multimodal multitask deep

learning," *Journal of Network and Computer Applications*, vol. 183, p. 102985, 2021.

[26] W. H. Asquith, *Univariate Distributional Analysis with L-moment Statistics using R*. PhD thesis, Texas Tech University, 2011.

[27] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in *ICISSp*, pp. 108–116, 2018.

[28] M. Sauter, ""LOIC Will Tear Us Apart" The Impact of Tool Design and Media Portrayals in the Success of Activist DDOS Attacks," *American Behavioral Scientist*, vol. 57, no. 7, pp. 983–1007, 2013.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

A protocol for data exchange with free samples using smart contracts

Rafael Genés-Durán, Juan Hernández-Serrano, Oscar Esparza, Miquel Soriano, José Luis Muñoz-Tapia
Departamento de Ingeniería Telemática,
Universitat Politècnica de Catalunya (UPC)
(rafael.genes,j.hernandez,oscar.esparza,miquel.soriano,jose.luis.munoz)@upc.edu

Marta Bellés-Muñoz
Universitat Pompeu Fabra (UPF)
belles.mm@gmail.com

Distrust between data providers and data consumers is one of the main obstacles hampering digital-data commerce to take off. Data providers want to get paid for what they offer, while data consumers want to know exactly what are they paying for before actually paying for it. In this paper, we summarize a protocol that overcomes this obstacle by building trust based on two main ideas. First, a probabilistic verification protocol, where some random samples of the real dataset are shown to buyers in order to allow them to make an assessment before committing any payment; and second a guaranteed, protected payment process, enforced with smart contracts on a public blockchain, that guarantees the payment of the data if and only if the data provided meets the agreed terms, and that refunds honest players otherwise.

***Palabras Clave*—data exchange, smart contract, blockchain, DLT, payments**

I. INTRODUCCIÓN

The use of data has increasingly become a crucial factor in the success of businesses. Businesses not only collect and analyse the data they generate, but increasingly rely on third party data to enhance its business value. In general, making proper data agreements is not easy, specially the task of valuing data and convincing customers of their value without giving them away [1]. The creation of marketplaces addresses many of these problems. Allowing providers and consumers to deal with common interests in a platform where both parties can meet each other and trade information solves the integration problem of connecting consumers and providers.

This article focuses on the problem of convincing consumers of data value, which can be seen as a form of lack of trust towards data providers. Traditionally, this problem cannot be solved without previously establishing confidence between parties. This represents an entry barrier to new providers in the market, hurting competence and thus, reducing utility for consumers. To exchange value safely, it is essential to ensure that consumers get the product they are paying for and that providers get paid. These two things are often carried out without any strict protocols and guaranteed just by existing trust. Typically, counterparties that know each other from previous experience or that are aligned with future interests, are confident that no intent to scam will be made by the other party, since confidence is often more beneficial than gains from fraud.

But, when stronger assurance than that is needed, it is a common practice to use a trusted third-party (TTP) to whom all parties trust to guarantee that the process is carried out correctly by all individuals involved. TTPs entail an extra cost for all parties, and generate a single point of failure that could produce critic delays and denial of services. Distributed Ledger Technologies (DLTs) can be seen as a paradigm shift when it comes to the need of TTPs. Using DLTs, all participants in the network can maintain a set of synchronized data (who owns what) without the need for a central authority (TTP) guaranteeing integrity, fairness and data availability.

In this paper we summarize DEFS (Data Exchange with Free Sample Protocol), a protocol that addresses the lack-of-trust between providers and consumers in a data trade. DEFS preserves the security, privacy and fairness standards that marketplaces should guarantee, and it includes the capability of checking some sample portions of the dataset before committing to purchase.

II. BACKGROUND

A. Merkle Hash Trees

A Merkle hash tree (MHT) is an authenticated data structure where every leaf node of the tree contains the

cryptographic hash of a data block and every non leaf node contains the concatenated hashes of its child nodes [2]. MHTs allow to link a set of data to a unique hash value, the Merkle hash tree root (MHR), allowing efficient and secure verification of consistency and content of large sets of data.

Figure 1 contains an example of a MHT with 8 leaves. To show that a certain value is stored in a leaf of the MHT, one can create a Merkle proof (MP), which consists of a list of the additional nodes required to compute the root of the tree. For instance, a Merkle proof showing that h_3 is stored in the MHT from Figure 2 would consist of the nodes h_2, h_{01} h_{4567}, $h_{01234567}$. Note that with h_3 and the first three nodes of this list anyone can compute the root of the tree. If the root matches $h_{01234567}$, then the proof is valid proof of membership for h_3 in the tree.

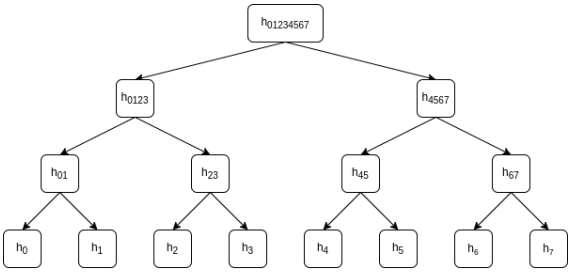


Fig. 1. MHT of 8 leafs.

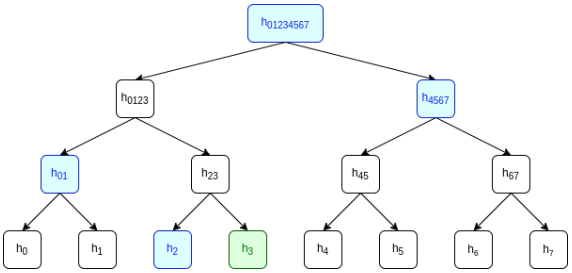


Fig. 2. Merkle proof for h_3. MP(h_3)=$h_{01234567}$,h_{4567},h_{01},h_2

III. State of the Art

Decentralized marketplaces have arisen as a solution to enhance security, sovereignty and trust in data exchanges [3], [4], [5].

One interesting initiative is GAIA-X [6], which is an European project to develop the foundations for a federated open data infrastructure connecting both classical architectures with decentralized infrastructures in order to build a transparent ecosystem for the end users taking advantage of the decentralized benefits.

One of the main technologies that is fostering data marketplaces is the Internet of Things (IoT) with huge amounts of data being generated from sensors and devices. The increasing necessity of monetizing these data is also pushing research. In the literature, we can find several works that propose decentralized marketplaces for IoT using distributed ledger technologies to enhance the data exchanges with transparency, trust and integrity [7], [8], [9]. Among others, decentralized marketplaces are being implemented in new disruptive scenarios such as artificial intelligence [10], smart cities [11], [12], and the connected car [13]. In fact, the value of the data is becoming more and more important to the business interactions which is reflected in the new technologies and their necessity to generate this new era of decentralized marketplaces.

An example of a decentralized data trading solution is presented in [14]. As in our protocol, the data on sale are not stored on the blockchain but in some external (and possibly distributed) storage platform. Similar to our protocol, the proposed solution symmetrically encrypts data on sale and uses a Merkle tree of cryptograms to register the associated trades on the blockchain. However, the solution proposed not only requires to generate symmetric cryptograms but also each of these cryptograms needs to be asymmetrically signed. Additionally, authors propose to use Plaintext Checkable Encryption (PCE) [15] to check on-chain that the cryptograms have been correctly encrypted. In DEFS, we avoid using asymmetric encryption, which is much slower than symmetric encryption.

Another remarkable implementation of a decentralized data trading solution is presented in [16], where authors present SDTE, a secure blockchain-based data trading ecosystem. As our protocol, SDTE tries to mitigate the existence of dishonest parties in data exchanges. However, SDTE focuses on an scenario in which the buyer does not need to have access to a complete dataset but it only needs the findings from the data analysis. For this case, SDTE proposes a data processing-as-a-service, where the buyer is paying for the analysis of the seller's dataset. SDTE is build using an Intel's SGX-based secure execution environment to protect the data processing, the source data and the analysis results. As we will show in the following section, DEFS is not designed as a data processing-as-a-service but as a data exchange-as-a-service. In the latter, the seller wants to buy the complete dataset not computed data. For this scenario, DEFS provides a probabilistic verification protocol and a conflict resolution protocol that is guaranteed and supported by a smart contract.

IV. Data Exchange Protocol

In this section we summarize DEFS, a protocol that addresses the problem of data trading between provider and consumers using a smart contract deployed in the blockchain as a broker. As we explained before, the use of DLTs can replace the role of TTPs in payment processes. When using DLTs, participants in the network can maintain synchronized data and share payment information without the need of a central authority, guaranteeing this way the integrity, fairness and availability of the data. In this manner, DEFS makes use of a smart contract to preserve the security and privacy standards that marketplaces should guarantee.

Another gap to cover in this data trading scenario is generating trust between data consumers and data providers. Here it comes the novelty of DEFS: our proposed data exchange protocol is designed with the capability of

checking random samples from the dataset, so that consumers are able to infer if the complete dataset is worth to be paid for, enhancing the trust of the consumer's side. On the other side, the smart contract acts as a broker during the payment procedure, ensuring providers that they will receive the payment for the data they exchanged.

A. Protocol Explanation

We assume that before starting the protocol, a data provider advertises her data to the public using off-blockchain means, such as a data marketplace. Then, a consumer interested in a particular dataset contacts the provider, who starts the DEFS protocol to perform the data exchange and payment. To prevent potential extensive leaks of the data, it is important that there is one DEFS protocol per each individual consumer. DEFS consists of three different phases:

1) **Protocol preparation**: in this initial phase, the provider prepares not only the data to be exchanged, but also all the parameters and cryptographic material necessary to demonstrate that the data exchange is secure and private. More specifically, the provider:
   - Divides the complete dataset in portions. These portions are chosen randomly from the dataset (not consecutively).
   - Generates a seed to generate symmetric cryptographic keys.
   - Uses these keys to create a MHT, whose root can be used to check the correctness of this cryptographic material.
   - Encrypts a random permutation of the data portions with the keys, obtaining an encrypted and randomized version of the whole dataset.
   - Creates another MHT using the hashes of these cryptograms as leaves, whose root can be used to verify the correctness of the cryptograms generated.
   - Deploys a smart contract in the blockchain containing certain public parameters and that smart contract acts as a broker during the rest of the protocol.

   If the consumer has interest in obtaining the dataset, the protocol continues as follows:
   - The consumer receives the whole dataset encrypted but it cannot be decrypted at that very moment.
   - The consumer queries the smart contract to obtain the root of the tree of cryptograms and verifies that all the cryptograms belong to this tree.

   At this point, all entities (consumer, provider and smart contract) are ready to start the protocol execution phase, in which the consumer will have access to the complete dataset and perform the payment.
2) **Protocol execution**: in this phase, the consumer will be able to get some samples of the dataset (for free) to evaluate if it is worth to pay, and if so, it will obtain the dataset and the provider will be paid:
   - The consumer will choose at random some sample portions to be revealed.
   - The provider will disclose the keys for those samples, so the consumer can evaluate the quality of the dataset.
   - If the consumer is not convinced, the protocol ends here. However, if it decides that it is worth paying the dataset, it will commit the payment to the smart contract.
   - The provider is asked to publish the seed (that will disclose all the encryption keys) in the smart contract.
   - If the consumer is able to properly decrypt the dataset, after a timeout, the provider is paid and the protocol ends.
   - If the consumer is able to prove that there were problems with the previous procedure, it starts the conflict resolution phase to obtain a refund.

   The following phase will only be needed in case the consumer considers that is cheated on.

3) **Conflict resolution***: this phase is optional, it only takes place if the consumer detects a provider misbehaviour. The following are the cases that can end with a refund if he is able to demonstrate this misbehaviour:
   - Keys are not properly generated.
   - Cryptograms do not have the proper format.

B. Protocol Properties

The main properties provided by our protocol are the following:

1) **Data samples evaluation**. The consumer gets a free set of fair samples of the data being traded before paying. The protocol ensures that neither the consumer nor the provider are able to manipulate the chosen data or select specific samples.
2) **Payment guarantees**. The provider gets paid if and only if the consumer has access to the whole set of data. That is, the consumer can not get the data without paying for it and the provider does not get paid without disclosing the data.
3) **The solution is cost-efficient**. Due to high fees on public ledgers, DEFS minimizes the amount of data stored on the network, which is also independent of the quantity of data traded. This way, both the amount of data stored and the number of interactions with the distributed ledger is constant.
4) **Non-repudiation**. The DEFS protocol ensures that any party involved in the exchange is not able to cancel and/or deny the data exchange once an agreement is made. Since the hash function used to generate the MHT is assumed to be collision-resistant, the MRC and MRK logged in the smart contract creation will prevent other data to be faked as bought or sold this way. Moreover, the use of a public blockchain enhances the integrity of the actor actions.

5) **Liveness**. The different timeouts guarantee that the protocol reaches a final state, even when one of the parties quits in advance. The provider can cancel the smart contract if no consumer reaches him out and the timeouts set after payments ensure that any counterparty can finalize the execution of the protocol favourably for it if the other party does not act on time.

C. State Diagram

The protocol operation and the interactions between the different stakeholders and the smart contract are detailed in Figure 3.

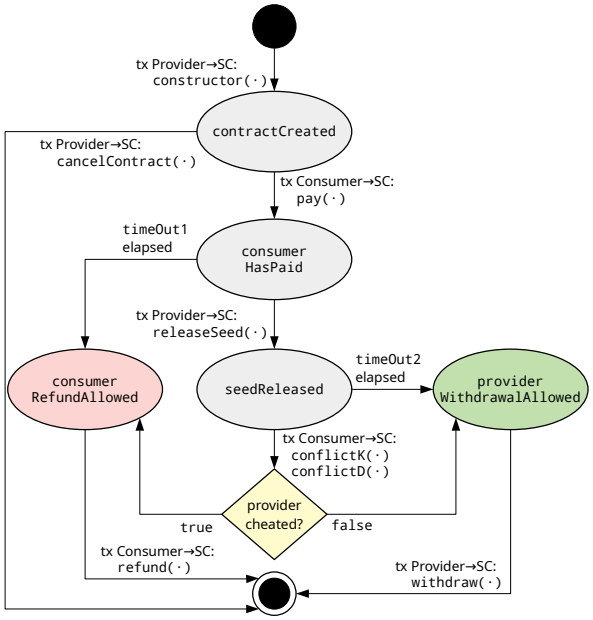


Fig. 3. State diagram of the smart contract.

V. Conclusions

Distrust is one of the main obstacles to implement exchanges between data providers and data consumers in a decentralized way. In this article, we summarize a protocol that allows a consumer to probabilistically obtain and check a subset of a dataset on sale from a provider before committing the payment. The protocol is executed using a smart contract deployed in a public distributed ledger. Once the consumer accepts to buy the dataset, the payment process, the agreed terms, and the possible refunds are managed and enforced by the smart contract. To expose the dataset, our protocol splits the data in portions and encrypts and stores each portion off-chain. Then, we create a MHT for the cryptograms and another MHT for the encryption keys. The encryption keys are related to each other using a cryptographic hash function in a way that allows us to implement a cost-efficient conflict resolution mechanism. The security analysis of our protocol shows that consumers and providers are economically protected and that the provider can reduce the risks of identity-replication attacks by adjusting the amount of free samples disclosed to the consumer.

Agradecimientos

This research has been funded by i3Market (H2020-ICT-2019-2 grant number 871754). This work is also supported by the TCO-RISEBLOCK (PID2019-110224RB-I00), ARPASAT (TEC2015-70197-R), Project RTI2018-102112-B-I00 (AEI/FEDER,UE) and by the Generalitat de Catalunya grant 2014-SGR-1504.

Referencias

[1] L. D. W. Thomas and A. Leiponen, "Big data commercialization," *IEEE Engineering Management Review*, vol. 44, no. 2, pp. 74–90, Second 2016.

[2] F. Haider, "Compact sparse merkle trees," Cryptology ePrint Archive, Report 2018/955, 2018, https://eprint.iacr.org/2018/955.

[3] H. Yoo and N. Ko, "Blockchain based data marketplace system," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 1255–1257.

[4] L. Mikkelsen, K. Mortensen, H. Rasmussen, H.-P. Schwefel, and T. Madsen, "Realization and evaluation of marketplace functionalities using ethereum blockchain," in *2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC)*, 2018, pp. 47–52.

[5] V. P. Ranganthan, R. Dantu, A. Paul, P. Mears, and K. Morozov, "A decentralized marketplace application on the ethereum blockchain," in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, 2018, pp. 90–97.

[6] A. Braud, G. Fromentoux, B. Radier, and O. Le Grand, "The road to european digital sovereignty with gaia-x and idsa," *IEEE Network*, vol. 35, no. 2, pp. 4–5, 2021.

[7] K. R. Ăzyilmaz, M. DoÄan, and A. Yurdakul, "Idmob: Iot data marketplace on blockchain," in *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, 2018, pp. 11–19.

[8] D.-D. Nguyen and M. I. Ali, "Enabling on-demand decentralized iot collectability marketplace using blockchain and crowdsensing," in *2019 Global IoT Summit (GIoTS)*, 2019, pp. 1–6.

[9] P. Tzianos, G. Pipelidis, and N. Tsiamitros, "Hermes: An open and transparent marketplace for iot sensor data over distributed ledgers," in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2019, pp. 167–170.

[10] V. Arya, S. Sen, and P. Kodeswaran, "Blockchain enabled trustless api marketplace," in *2020 International Conference on COMmunication Systems NETworkS (COMSNETS)*, 2020, pp. 731–735.

[11] S. Musso, G. Perboli, M. Rosano, and A. Manfredi, "A decentralized marketplace for m2m economy for smart cities," in *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2019, pp. 27–30.

[12] G. S. Ramachandran, R. Radhakrishnan, and B. Krishnamachari, "Towards a decentralized data marketplace for smart cities," in *2018 IEEE International Smart Cities Conference (ISC2)*, 2018, pp. 1–8.

[13] B.-G. Jeong, T.-Y. Youn, N.-S. Jho, and S. U. Shin, "Blockchain-based data sharing and trading model for the connected car," *Sensors*, vol. 20, no. 11, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/11/3141

[14] Y.-N. Li, X. Feng, J. Xie, H. Feng, Z. Guan, and Q. Wu, "A decentralized and secure blockchain platform for open fair data trading," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 7, p. e5578, 2020, e5578 cpe.5578. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5578

[15] S. Ma, Y. Mu, and W. Susilo, "A generic scheme of plaintext-checkable database encryption," *Information Sciences*, vol. 429, pp. 88–101, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025517301640

[16] W. Dai, C. Dai, K.-K. R. Choo, C. Cui, D. Zou, and H. Jin, "Sdte: A secure blockchain-based data trading ecosystem," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 725–737, 2020.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Modelado del Conocimiento de Ciberseguridad en Entornos Hospitalarios

Susel Fernandez, Luis Cruz-Piris, Ivan Marsa-Maestre, Jose Manuel Gimenez-Guzman.
Departamento de Automática,
Universidad de Alcalá
Escuela Politécnica Superior. Campus Universitario, Ctra. Madrid-Barcelona km. 33, 600. 28805. Alcalá de Henares. Madrid.
susel.fernandez@uah.es, luis.cruz@uah.es, ivan.marsa@uah.es, josem.gimenez@uah.es.

En la actualidad se está produciendo un considerable incremento en el número de ataques informáticos dirigidos a infraestructuras críticas en todo el mundo, entre las que destacan las infraestructuras de los centros hospitalarios. Estos ataques pueden tener graves consecuencias, desde el filtrado masivo de datos sanitarios confidenciales, hasta el colapso total de las infraestructuras TICs de las entidades sanitaras. Este trabajo se centra en la gestión del conocimiento de ciberseguridad en el entorno hospitalario y consiste en el desarrollo de una ontología que modele los principales conceptos y relaciones identificados en este dominio. El objetivo principal es poder contar con infraestructuras capaces de detectar estos ataques y establecer mecanismos de actuación que permitan mitigar sus efectos sobre los activos en un entorno tan crucial y vulnerable como es el entorno sanitario.

***Palabras Clave*- gestión de conocimiento, ontología, ciberseguridad.**

I. INTRODUCCIÓN

En los últimos años estamos asistiendo a un crecimiento en el número y tipología de las amenazas de seguridad, que afectan no sólo a los ordenadores de los usuarios, sino a todo tipo de dispositivos y sistemas. Se está produciendo un considerable incremento en el número de ataques informáticos dirigidos a infraestructuras críticas de todo tipo en todo el mundo. Entre estos destacan especialmente los ciberataques sobre las TICs en centros de atención hospitalaria [1]. La naturaleza de estos incidentes es cada vez más variada y compleja y va desde simples ataques de denegación de servicio (DoS) hasta infecciones por malware, como por ejemplo los ataques de tipo Ransomware [2] [3], donde los atacantes interrumpen las operaciones informáticas y solicitan dinero para "liberar" los recursos bloqueados. Los motivos para realizar estos ataques se basan fundamentalmente en la falta de seguridad de las instalaciones hospitalarias y en el gran impacto que puede tener la falta de acceso a los recursos informáticos en estas instituciones.

En mayo del 2017, una fracción significativa de hospitales en el Reino Unido fue atacada por un malware, que provocó una considerable afectación a las áreas de emergencias y retrasos en las cirugías [4]. Más recientemente, en septiembre de 2019, el Campbell County Memorial Hospital en Wyoming, EE. UU., fue golpeado por un ataque de malware más específico, que provocó la cancelación de muchos servicios como las cirugías y los exámenes de radiología.

En España, en el mes de enero de 2020, el Hospital Universitario de Torrejón, quedaba paralizado por un ataque informático de tipo Ransomware. Casi dos semanas después, el centro sanitario conseguía rescatar su sistema de citas automatizadas, pero otros tantos servicios permanecieron inactivos por más tiempo. El incidente colapsó por completo el sistema informático del hospital [5].

Esta creciente proliferación de incidentes de seguridad se debe a varios factores, entre los cuales se encuentra la conectividad de los diferentes dispositivos de las instituciones hospitalarias a las redes y por supuesto, a internet. Dentro de una red de atención médica, podemos encontrar en la actualidad una gran variedad de dispositivos, que van desde ordenadores portátiles, smartphones, hasta equipos médicos específicos, que generan y procesan datos médicos. Esta combinación de dispositivos heterogéneos junto con la ubicuidad deseada y la alta conectividad [6] implica un desafío complejo desde el punto de vista de la ciberseguridad.

Lo deseable sería poder contar con una infraestructura diseñada para reaccionar ante incidentes de seguridad, que limite la propagación del ataque y permita mitigar sus efectos de manera rápida y eficaz. Para lograr este objetivo, la gestión del conocimiento de seguridad en el entorno de las TICs en los centros hospitalarios es crucial.

Este trabajo está dirigido a abordar la necesidad de contar con un modelo del conocimiento lo suficientemente expresivo, que permita modelar las diferentes características de los ciberataques más comunes, sus relaciones con las vulnerabilidades y cómo afectan a los activos identificados en este dominio. Para el modelado de la base de conocimiento del sistema se ha desarrollado una ontología, que abarca los principales conceptos y relaciones del dominio de la ciberseguridad en el entorno hospitalario, así como una base de reglas para el mecanismo de razonamiento con la misma.

El documento está organizado de la siguiente manera. La sección II presenta los antecedentes y el estado actual del tema. En la sección III se presenta el modelado del conocimiento, que incluye la ontología y el mecanismo de razonamiento. En la sección IV se describen brevemente algunos escenarios de prueba y finalmente, las conclusiones y líneas de trabajo futuro se resumen en la sección V.

II. Antecedentes y estado actual del tema

En el área de la ciberseguridad existen algunas bases de datos con información sobre vulnerabilidades, productos y componentes a los que afectan, especificaciones técnicas, impacto y vector de ataque. También hay herramientas para el análisis automático de vulnerabilidades, basadas en configuraciones recomendadas de seguridad (checklists) para sistemas y servicios, que van asociadas al nivel de protección necesario en función del tipo de sistema.

A. Bases de datos y clasificación de vulnerabilidades

En cuanto a las bases de datos de vulnerabilidades tenemos que mencionar a CVE (Common Vulnerabilities and Exposures) [7] y CWE (Common Weakness Enumeration) [8], ambas de MITRE Corporation. CVE, un diccionario público que proporciona un identificador único para cada vulnerabilidad, se ha convertido en un estándar y constituye la principal fuente de información sobre vulnerabilidades para otras bases de datos. Por su parte, CWE también es un estándar internacional y de libre uso, que proporciona un lenguaje común para describir vulnerabilidades de seguridad de software en arquitectura, diseño y codificación. Otra base de datos conocida es la del gobierno estadounidense, NVD (National Vulnerability Database) [9], que permite la automatización de la gestión de vulnerabilidades y la medición del nivel de seguridad. Incluye listas de comprobación de configuraciones de seguridad de productos y efectos en software relacionado.

Un elemento importante a la hora de abordar las vulnerabilidades es poder clasificarlas según su severidad para establecer estrategias efectivas de protección. CVSS (Common Vulnerability Scoring System) [10] es un sistema que permite calcular la severidad de una vulnerabilidad, de manera estricta a través de fórmulas matemáticas. Proporciona un estándar para comunicar las características y el impacto de una vulnerabilidad identificada con su código CVE.

En cuanto a la clasificación de vulnerabilidades, existe OWASP (Open Web Application Security Project), que es un proyecto de código abierto dedicado a determinar y combatir las vulnerabilidades en aplicaciones en el entorno Web. OWASP Top Ten [11] es un documento que se publica cada tres años, con los diez riesgos de seguridad más importantes en aplicaciones Web. Su objetivo es crear conciencia acerca de la seguridad en aplicaciones mediante la identificación de los riesgos más críticos que enfrentan las organizaciones.

B. Herramientas de análisis de vulnerabilidades

En cuanto a las principales herramientas disponibles, tenemos las que permiten realizar análisis de vulnerabilidades de código fuente. RATS (Rough Auditing Tool for Security) [12] es un analizador de código estático, *open source*, para detectar potenciales problemas de seguridad en varios lenguajes de programación como C/C++, Perl, PHP, Python y Ruby. Entre las herramientas para el análisis de vulnerabilidades de sistemas completos, tenemos MBSA (Microsoft Baseline Security Analyzer) [13], que permite analizar la seguridad de pequeñas redes formadas por equipos con Windows. Entre los escáneres de vulnerabilidades más completos tenemos OpenVAS (Open Vulnerability Assessment Scanner) [14], una herramienta *open source*, que abarca un espectro amplio de funciones, incluyendo varios protocolos industriales, escaneos a gran escala y un potente lenguaje de programación para implementar pruebas de vulnerabilidad. Por su parte, MITRE ATT&CK [15] es una base de conocimiento accesible a nivel mundial, para el desarrollo de modelos y metodologías de amenazas específicas de ciberseguridad en varios dominios.

C. Ontologías

Existen algunos modelos ontológicos dirigidos a abordar distintas áreas de la ciberseguridad. En [16] proponen una ontología de seguridad utilizando lógica descriptiva, destinada a organizar el conocimiento relacionado con la gestión de riesgos. En [17] se presenta una ontología diseñada para el análisis y la gestión de vulnerabilidades, que incluye las relaciones entre productos TIC, vulnerabilidades, atacantes, métricas de seguridad y contramedidas. En [18] introducen una ontología que captura información sensible a la privacidad para los Sistemas de Redes Sociales. La ontología permite detectar ausencia de políticas de protección de privacidad. Otras ontologías se centran en ataques específicos, como las amenazas persistentes avanzadas (ATP), que pueden materializarse a través de distintos tipos de malware en diversos entornos, como por ejemplo en dispositivos IoT [19, 20]

Aunque ha habido varias propuestas de ontologías en el estado del arte, la mayoría están en las primeras etapas de desarrollo y su reutilización para nuestro propósito es bastante limitada. Ninguna de las aproximaciones existentes ha demostrado ser lo suficientemente expresiva para cubrir el conocimiento necesario para responder de manera adecuada y rápida ante los ciberataques en entornos hospitalarios.

III. Modelado del conocimiento

Para el modelado del conocimiento se ha diseñado una ontología, que abarca los principales conceptos y sus relaciones en el entorno de la ciberseguridad en infraestructuras hospitalarias.

A. *Ontología.*

La ontología ha sido desarrollada en lenguaje OWL [21], utilizando la herramienta *Protégé* [22], que proporciona una interfaz gráfica interactiva y amigable para el trabajo con estas estructuras. La ontología diseñada se compone de tres bloques de conocimiento fundamentales: vulnerabilidades, amenazas y activos del sistema que pueden verse afectados por la materialización de las amenazas. La fig. 1, presenta una parte de la ontología diseñada, en la que se pueden apreciar los principales conceptos modelados en cada bloque de conocimiento y sus relaciones.

Como se puede observar en la figura, entre los principales activos del sistema, susceptibles a incidentes de seguridad se han identificado los sistemas de cuidado remoto, los dispositivos médicos en red, los sistemas de identificación, el equipamiento de red, los sistemas de información clínica interconectados, los datos y las instalaciones, entre otros.

En el bloque de conocimiento de las amenazas se recogen los principales ataques a los que suelen estar sometidos estos sistemas, entre los que destacan los ataques de robo de identidad, el *phishing*, la manipulación no autorizada de los dispositivos médicos, el secuestro de sesiones, el robo de datos, la denegación de servicio (DoS) y las infecciones por malware, que pueden llegar a ser muy variadas en cuanto a características, modos de infección y consecuencias sobre el sistema.

El otro bloque de conocimiento presentado es el que permite relacionar los ataques con los activos del sistema: las vulnerabilidades. En otras palabras, las vulnerabilidades pueden definirse como las debilidades que tiene el sistema que hacen posible que las amenazas se puedan materializar, causando daños sobre los activos. Las principales vulnerabilidades se han agrupado en dos categorías. La primera, vulnerabilidades de diseño e implementación, agrupa los conceptos relacionados con las debilidades del sistema provocadas por fallos en el desarrollo de las aplicaciones, ya sea en fase de diseño o de implementación, como por ejemplo el buffer overflow, el Cross-Site Scripting (XSS) o la inyección SQL, entre otros. La segunda categoría, se centra en los conceptos relacionados con las vulnerabilidades de configuración, que pueden ser por ejemplo un mal uso de los sistemas de autenticación, el uso de configuraciones predeterminadas, las cuentas de usuario no seguras, el almacenamiento de la información no cifrada, entre otros.

B. *Razonamiento*

El mecanismo de razonamiento es crucial cuando se trabaja con ontologías, porque es el proceso que permite realizar la inferencia de nuevo conocimiento a partir del conocimiento ya existente en la ontología.

En este trabajo hemos utilizado el razonador semántico *Pellet* [23], que permite además validar la consistencia de la ontología. Las reglas de razonamiento en la ontología propuesta, se han desarrollado utilizando el Lenguaje de Reglas de la Web Semántica SWRL [24]. Una vez poblada la ontología, con los individuos correspondientes a los diferentes escenarios de seguridad, la base de reglas permite realizar consultas para testar el nivel de expresividad de la ontología.

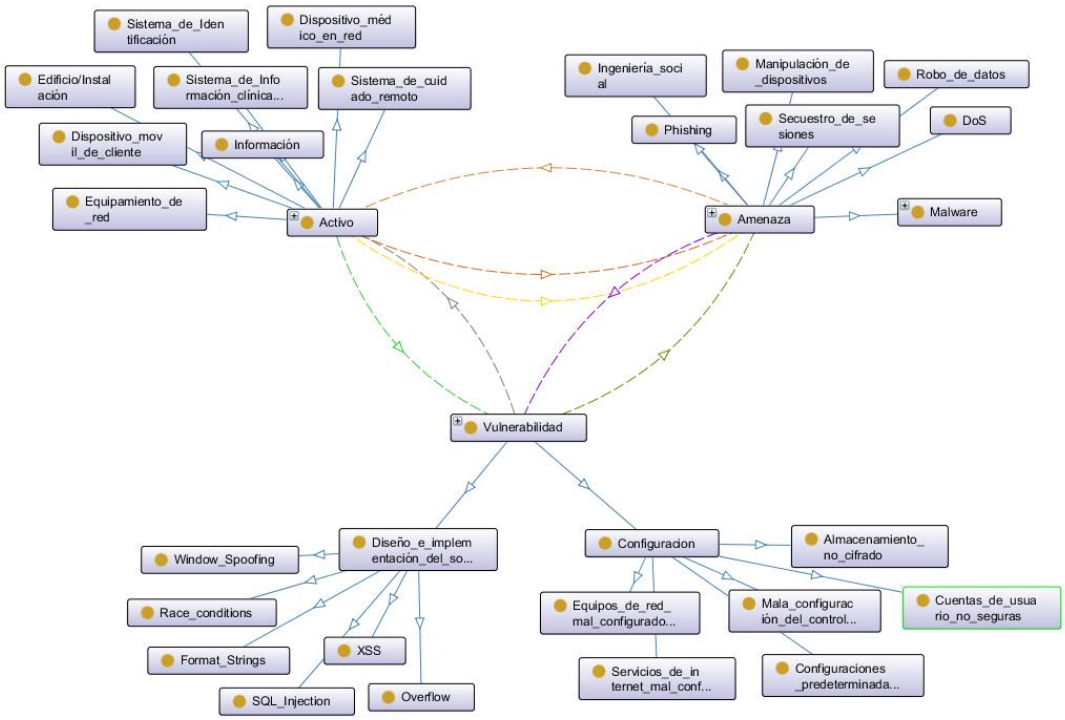


Fig.1 Ontología de ciberseguridad en infraestructuras hospitalarias

IV. Pruebas de expresividad de la ontología

El trabajo se encuentra actualmente en la fase de pruebas de la expresividad de la ontología. Para la realización de las pruebas se han definido varios escenarios que permiten obtener información de la ontología para prevenir ataques típicos a las TICs en infraestructuras hospitalarias. Las consultas a la ontología se han diseñado utilizando el lenguaje SQWRL [25]. Entre estas consultas se encuentran los listados o reportes de distintos tipos de amenazas y las vulnerabilidades que explotan, como por ejemplo:

- Amenazas de mala reputación en datos clasificados.
- Amenazas por errores involuntarios de usuarios.
- Amenazas de distribución de software malicioso por parte de los usuarios.
- Amenazas de ingeniería social.
- Malwares que utilizan una técnica especifica de ataque.
- Amenazas registrados sobre dispositivos médicos específicos.
- Amenazas que explotan vulnerabilidades específicas.
- Dispositivos con vulnerabilidades específicas conocidas.
- Listado de dependencias HW y SW para poder llevar a cabo el análisis de riesgos.

También se han definido consultas sobre políticas de seguridad tanto generales como específicas de distintos sistemas y dispositivo, permitiendo realizar un filtrado por distintas categorías.

V. Conclusiones

En este trabajo se presenta una ontología para la gestión del conocimiento de ciberseguridad en infraestructuras hospitalarias. El objetivo principal es proporcionar un modelo semántico lo suficientemente expresivo, que abarque las diferentes características de los ciberataques más comunes, sus relaciones con las vulnerabilidades y cómo afectan a los activos principales de este dominio. Esto permitirá contar con una infraestructura resiliente, que limite la propagación de cualquier ataque y facilite la mitigación de sus efectos de manera rápida y eficaz.

La expresividad de la ontología se encuentra actualmente en fase de pruebas. Para ello se han definido una serie de consultas en diferentes escenarios de ciberseguridad. Como trabajo futuro se pretende continuar mejorando la expresividad de la ontología, añadiendo más conceptos, relaciones y reglas de razonamiento y desarrollar una interfaz gráfica que facilite la gestión del conocimiento modelado de una manera más amigable.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto ESCUDO- Sistema de gEStión del conocimiento de CibersegUridaD en entornos hOspitalarios. CCG20/IA-041, de la Universidad de Alcalá y el proyecto CloudWall-Cloud-enabled Resilience Framework PID2019-104855RBI00/AEI/10.13039/501100011033 del Ministerio de Ciencia e Innovación.

Referencias

[1] Safavi, S., Meer, A. M., Melanie, E. K. J., & Shukur, Z. (2018, November). Cyber Vulnerabilities on Smart Healthcare, Review and Solutions. In 2018 Cyber Resilience Conference (CRC) (pp. 1-5). IEEE.

[2] Verizon 2019 Data Breach Investigations Report. Available at https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf. Accedido el 10 de marzo de 2020.

[3] ENISA. Security and Resilience for Smart Health Service and Infrastructures. 2016. Available at https://www.enisa.europa.eu/publications/cyber-security-and-resilience-for-smart-hospitals. Accedido el 10 de marzo de 2020.

[4] Smart, W. (2018). Lessons learned review of the WannaCry ransomware cyber attack. London: Skipton House.

[5] Noticia. El diario. https://www.eldiario.es/tecnologia/Desterrar-solucion-ciberataques-hospital-Madrid_0_990051221.html. Accedido el 10 de marzo de 2020.

[6] T. Walker, Interoperability a must for hospitals, but it comes with risks, Manag. Healthc. Exec. (2017) http://managedhealthcareexecutive.modernmedicine.com/.

[7] Common Vulnerabilities and Exposures. https://cve.mitre.org/. Accedido el 10 de marzo de 2020.

[8] Common Weakness Enumeration. https://cwe.mitre.org/. Accedido el 10 de marzo de 2020.

[9] National Vulnerability Database. https://nvd.nist.gov/. Accedido el 10 de marzo de 2020.

[10] Common Vulnerability Scoring System (CVSS) https://nvd.nist.gov/cvss.cfm. Accedido el 10 de marzo de 2020.

[11] Open Web Application Security Project. Top ten Web Application Security Risks. https://owasp.org/www-project-top-ten/. Accedido el 10 de marzo de 2020.

[12] Rough Auditing Tool for Security (RATS). https://security.web.cern.ch/security/recommendations/en/codetools/rats.shtml. Accedido el 10 de marzo de 2020.

[13] Microsoft Baseline Security Analyzer (MBSA). https://www.microsoft.com/en-us/download/search.aspx?q=MBSA. Accedido el 10 de marzo de 2020.

[14] Open Vulnerability Assessment Scanner (OpenVas). https://www.openvas.org/. Accedido el 10 de marzo de 2020.

[15] Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., & Thomas, C. B. (2018). Mitre att&ck: Design and philosophy. Technical report.

[16] Fenz, S. and Ekelhart, A. (2009) Formalizing Information Security Knowledge. Proc. 4th Int. Symp. Information, Computer, and Communications Security, Sydney, Australia, March 10–12, pp. 183–194. ACM, New York, USA.

[17] Wang, J.A. and Guo, M. (2009) OVM: An Ontology for Vulnerability Management. Proc. 5th Annual Workshop on Cyber Security and Information Intelligence Research, Tennessee, USA, January 8–10, pp. 1–4. ACM, New York, USA.

[18] Masoumzadeh, A. and Joshi, J. (2013) Privacy Settings in Social Networking Systems:What You Cannot Control. Proc. 8th ACM SIGSAC Symp. Information, Computer and Communications Security, Hangzhou, China, May 8–10, pp. 149–154. ACM, New York, USA.

[19] Han, W., Xue, J., Wang, Y., Zhang, F., & Gao, X. (2021). APTMalInsight: Identify and cognize APT malware based on system call information and ontology knowledge framework. Information Sciences, 546, 633-664.

[20] Liang, X., Ma, L., An, N., Jiang, D., Li, C., Chen, X., & Zhao, L. (2019, December). Ontology Based Security Risk Model for Power Terminal Equipment. In 2019 12th International Symposium on Computational Intelligence and Design (ISCID) (pp. 212-216). IEEE.

[21] M. Dean, and G. Schreiber, OWL Web Ontology Language Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210/; 2004.

[22] Protégé: http://protege.stanford.edu/

[23] Pellet http://clarkparsia.com/pellet/

[24] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, Submission to W3C, May 2004 http://www.w3.org/Submission/SWRL/

[25] O'Connor, M. J., & Das, A. K. (2009, October). SQWRL: a query language for OWL. In OWLED (Vol. 529, No. 2009).


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Federated learning for smart charging of connected electric vehicles

Yaqoob Al-Zuhairi, Prashanth Kannan, Mónica Aguilar Igartua
Department of Network Engineering
Universitat Politécnica de Catalunya (UPC), Barcelona, Spain
yaqoob.al-zuhairi@upc.edu, prasuka4@gmail.com, monica.aguilar@upc.edu

Due to rising concerns over climate change, air pollution and clean energy awareness, the demand for electric vehicles (EVs) and renewable energy generation has increased in recent years. The main objective of this research is to design a decentralized smart charging coordination framework for EVs based on federated learning (FL) algorithms in order to provide an acceptable collaboratively learning model with privacy preservation of EVs, improve charging scenarios, contribute to smart grid stabilization, meet EVs energy requirements wherever and whenever they request, and gain welfare for EV owners. Moreover, FL is introduced with the goal of bringing machine learning (ML) down to the edge level in vehicular networks. Ultimately, a multimetric routing protocol is also used to predict the best route for transmitting messages among EVs, infrastructures, charging stations (CSs), and central servers.

***Key words*—FL, Edge Computing, VANET, V2X, Multimetric Routing Protocol**

I. INTRODUCTION

In recent years, the transport industry accounts for the bulk of greenhouse gas emissions and pollution to the environment [1]. To tackle this issue, governments are implementing policies to support use of renewable energies, decrease the dependency on crude oil, reduce CO_2 and pollutant emissions, and promote transition to more sustainable mobility. Current electric vehicles (EVs) as a part of the intelligent transport systems (ITS) have noticeably attracted substantial attention recently because they are very friendly to the environment as well as pollution-free vehicles. The recent prominent progress in the construction of charging infrastructure accelerates the penetration of EVs in the market. In 2018, the global EV fleet in the world exceeded 5.1 million units and it is expected to rise to 250 million units by 2030 [2].

Commercialization of the fifth-generation (5G) communication technologies and the emergence of vehicular networks and edge computing can promote a superior performance of the charging management [3]. In the meantime, future intelligent vehicles, which are at the heart of high mobility networks, are increasingly equipped with a wide variety of sensors to help the vehicle perceive the surrounding environment as well as monitor its own operational status in real time. Together with high performance computing and storage devices, these sensing technologies are transforming vehicles from a simple transportation facility to a powerful computing and networking hub with intelligent processing capabilities [4]. However, privacy concerns arise from the exchange of data with different parties. Privacy, in recent years, has been one of the most important concerns in vehicular environments. Thus, EVs may fail to exchange data among themselves and other parties due to privacy restrictions in situations where drivers are unwilling to provide their personal data due to the risk of data misuse and leakage [5]. In order to solve the aforementioned issue, federated learning (FL), which is a technique that enables distributed vehicles to collaboratively learn a shared machine learning (ML) model without sharing their raw data, would be a good solution to preserve the privacy of the local data [6].

Vehicular networks, e.g., vehicular ad hoc networks (VANETs) and cellular V2X (C-V2X), have gained popularity in recent years. In such networks vehicles equipped with wireless communication devices form vehicular networks [7]. However, with the evolution of technology and sudden growth in the number of smart vehicles, this will result in unprecedented pressure on communication infrastructures. Moreover, traditional VANET faces several technical challenges in deployment and management due to less flexibility, less resources, scalability, poor connectivity, and inadequate intelligence. Merging vehicular networks with edge computing, an emerging paradigm which moves computing tasks and services from the core to the network edge, i.e. closer to end users, is an appropriate solution for these types of challenging networks. Vehicular networks show special features such as high vehicles' mobility, difficult network connectivity, which is specially challenging for real-time applications

requiring low latency. The notion of exploiting vehicles as infrastructures could make the best use of several unused resources of vehicles to meet the ever increasing requirement in communication and computational capabilities [8]. FL can utilize vehicular big data generated from a large number of vehicles, and build a global ML model that could be used by any vehicle. FL can also be easily incorporated with edge computing where the edge vehicles provide an underlying infrastructure for FL [5][9].

Vehicular communications are crucial for exchanging data either between vehicles through vehicle-to-vehicle (V2V) or via vehicle-to-infrastructure (V2I) connections [10]. Although road side units (RSUs) enlarge the network communication capacity, they are really expensive and difficult to fully deploy along roads, particularly on a large scale such as over a whole city. Furthermore, future vehicles will need to communicate with everything around them in what is known as vehicle-to-everything (V2X) [10]. To enable hybrid vehicular communications, dedicated short-range communication (DSRC), which is based on IEEE 802.11p, and cellular V2X (C-V2X) can be adopted.

In this paper, we present the topics that will be studied in the first authors' doctoral thesis. In section II we highlight some related works. Section III summaries the basics of FL framework for vehicular networks. Since we consider the presence of EVs in our proposals, section IV depicts innovative ways to charge EVs, which will be taken into account in the design of our FL framework. Section V lists our ongoing work, whereas section VI shows which simulators are we going to use to evaluate the performance of our proposals. Finally, section VII concludes the paper.

II. LITERATURE SURVEY

The authors in [11] propose a scheme for charging moving EVs in a wireless fashion. An unsupervised ML algorithm is used to estimate the current charging status of each EV. Moreover, a routing protocol using distance vector information was used to advertise participating EVs about their state of charge (SoC), vehicle ID, distance information, and location. Authors mentioned that sharing such information among EVs causes concerns related to privacy reduction. Results in this work show to be reliable in terms of dynamic wireless charging in both static and dynamic scenarios.

In [9], a survey of technical challenges, possible solutions, open problems, and future research directions for applying FL in vehicular networks are discussed.

In [6], an FL approach for learning on edge devices is studied. The paper analyzes the effect of participating clients and the importance of client selection strategies in FL models. The authors investigate the performance of the FL model focusing on the effect of various parameters on its accuracy and training time, as well as comparing the performance to a traditional ML technique.

In the paper [12], a new proposal of a novel probabilistic multimetric routing protocol is presented. The proposal takes better forwarding decisions that guarantee the packet delivering to destination with the highest probability, while keeping the average packet delay low. Four designed metrics are considered in this article (distance to destination, vehicles' density, position of vehicle, and available bandwidth). OMNeT++, VEINS, and SUMO are used to conduct simulations in a realistic urban scenario.

III. IMPLEMENTATION OF AN FL FRAMEWORK

Instead of sending raw data to a central server, which is common in the traditional centralized ML approach, FL can be beneficial in vehicular environments to support cooperation of multiple EVs with the central server. This way, distributed EVs will train a partial model using own local data to ensure privacy protection. The global model in the central server will be updated from the parameters' aggregation of those partial models.

The basic steps for an FL framework are as follows:

- Client selection: Initially, the central server must specify those EVs that should be involved in the model training.
- Model dissemination: Once the EVs are selected, the central server broadcasts an initial learning model for the training to the selected EVs.
- Distributed learning: Each EV trains the model based on its own local dataset, and then calculates its local updates of the global model.
- Global model aggregation: After a predefined training period to aggregate the new version of the global model, all EVs send only their updates of the common global model to either the central server or to an aggregator. An EV can be designated as the aggregator looking to be close to participating EVs. The aggregator EV will manage the training process and ultimately send the result to the central server [13].
- Model testing: The central server tests the aggregated global model. According to the testing results, the central server could tune some hyper-parameters to repeat the training process, or continue with the next step which is the model update.
- Model update: The server updates the shared model according to the aggregated result from the EVs. Then, the server sends the updated global model to the EVs.

These steps are repeated until the central server achieves a satisfactory global model. While designing an FL framework, it is significant to select a variety of parameters and compare multiple scenarios to attain the trade-offs between privacy, efficiency, and accuracy in this environment [5]. Furthermore, different from conventional decentralized ML approaches, FL in vehicular environments is expected to achieve many advantages, such as low communication overhead, better privacy, larger data for training, better efficiency, better utility, and shorter response time.

IV. FUTURE WAYS TO CHARGE ELECTRIC VEHICLES

Currently, the most common form of charging EVs is via a plug-in charging station. However, there are many

issues reducing drivers' willingness to use EVs instead of traditional vehicles, such as limited number of charging stations (CSs). Furthermore, the nearest CS is perhaps at an inconvenient site in relation to an EV's route. In this point, the driver may be feeling anxiety over the need to recharge the EV during a journey. A further issue may happen after the EV reaches the plug-in CS, the driver may find the charging slots already taken by other EVs. These issues can effectively be solved by adopting a new structure of wireless charging technology for EVs based on the magnetic resonant coupling wireless power transfer technique, which was introduced in the last decade [11]. This technique is used to exchange charge among paired EVs in a static position as well as in dynamic position (motion). This structure can work together with plug-in charge EVs or operate independently in order to increase charging opportunities for EVs.

There are mainly two ways for wireless charging: [14]

- **Inductive charging:** It includes a transmitter coil, which is embedded in the floor of the charging area and connected to the power supply (grid), while a receiver coil which is embedded in the EV's chassis and connected to the EV's battery.
- **Wireless vehicle-to-vehicle (V2V) charging:** It is based on wireless power transfer technique which has high power transfer efficiency with a long transmitting distance. This technology can achieve charging between moving EVs.

In our work, we plan to use several charging technologies to meet requirements for EVs with a low charging state. We will design a FL-based framework to assists the driver in finding a suitable power source based on the current circumstances of the environment, e.g. energy requirement, distance to destination and to possible CSs, pricing, waiting time, ways of charging such as (mobile CS, plug-in CS, V2V charging, or inductive charging), and other factors. Our FL-based framework will be assisted by V2X communication, SoC detection, multimetric routing protocol, and edge computing. Figure (1) demonstrates the scenario of decentralized smart charging coordination for EVs based on FL.

This charging coordination can offer solutions for many problems, such as:

1) Offering solutions for the issue of having a limited number of CSs.
2) Solving the problem of increasing numbers of EVs in a plug-in EV network.
3) Reducing charging delay, and minimizing overall EV energy costs.
4) Avoiding the grid congestion, especially in the evenings, that results from simultaneously charging too many EVs.
5) Forecasting of EVs energy requirements.
6) Exchanging data among EVs or with the central server while ensuring EV's privacy preservation.
7) Generating monetary benefits for EV owners.
8) Finding the cheapest charging options by giving an EV driver the optimal choice for charging.

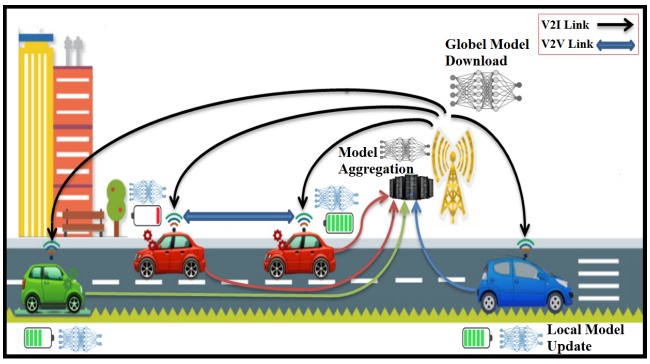


Fig. 1. Decentralized smart charging coordination framework for EVs based on federated learning (FL) algorithms.

9) Encouraging consumers to purchase EVs.

V. ONGOING WORK

The main contributions that we have planned in this Ph.D. thesis, can be summarized as follows:

1) Employing mobile EVs as infrastructures for communication and computation to achieve a better utilization of edge resources.
2) Privacy of each EV can be preserved by using FL.
3) Efficient collaboration amongst EVs with FL could reach the level of collaborative intelligence. Such intelligent collaboration can contribute to remarkably minimize the waiting time of charging.
4) Due to the potentially high mobility and sparse/dense environments, a multimetric routing protocol with FL can be employed to address the issue of providing an efficient forwarding scheme in V2X communications.
5) SoC, which indicates the level of charge available in each EV, is used to predict the residual energy of EVs based on the designed FL framework.
6) As for charging time available, determine the best power source either via wireless or plug-in through:
   - Considering the charging cost and duration when selecting either the best (e.g. the nearest) CSs or those EVs having high level of power to share, among other strategies.
   - Using an efficient multimetric routing protocol to improve the data transmission needed in the charging service regarding the chosen power source.
7) Estimating travel time to save energy according to the charging available time and the EV's destination, and boost the energy economy of EVs in charging processes.
8) Based on our FL-based charging framework, reducing load on CSs while keeping the grid stable can be achieved.

VI. IMPLEMENTATION AND SIMULATION

Because of the complexity and the high deployment costs of vehicular applications, it is recommended that

proposals are extensively tested and evaluated in possible ways before being put into practice.

TensorFlow [15] is an open-source FL library that offers an extensive range of algorithms. It also works in conjunction with Keras and Python. The construction of ML models and the design of parameters can be evaluated by computing metrics such as accuracy or confusion matrix.

Through simulations we will represent the whole system trustworthy as close to reality as possible, and we will assess the performance of the proposed system.

To simulate a vehicular network in urban scenarios, OMNeT++, SUMO, VEINS, and Artery-C are used.

- **OMNeT++:** It is a modular, object oriented and discrete event simulator based on C++ [16].
- **Simulation of Urban Mobility (SUMO) tool:** It is a highly portable and open-source software to simulate the movement of the vehicles [17].
- **Vehicles in networks simulations (VEINS) tool:** It is a framework for vehicular network simulation. It facilitates the bidirectional interaction among SUMO and OMNeT++ simulators [18].
- **Artery-C framework:** Artery-C [19] is a simulation framework for the performance evaluation of Cellular V2X protocols and V2X applications. It is an extension of the simulation framework SimuLTE [20], developed under the OMNeT++ platform.

VII. CONCLUSIONS

In vehicular environments, it is not always feasible to transfer massive amounts of data for processing to a remote central server due to limiting factors, such as unstable and limited wireless connectivity, unacceptable latency, and network bandwidth. Moreover, the interests on data availability have triggered the discussion of data ownership, and the concern on data privacy and confidentiality especially when sharing such data with central servers. In this research work we will design a decentralized smart charging coordination framework for EVs based on FL taking advantage of the distributed power of edge computing. Our goal is to encourage collaboration amongst participating EVs in training a global model locally while ensuring privacy preservation for each EV, and then sending it to a central server to obtain a global model. The aim is to optimize charging decisions, offer efficient EV charging service for urban areas when EVs are in a critical SoC, avoid grid overload, and acquire drivers' comfort and satisfaction. As a future work, an incentive mechanism will be used to motivate EVs with their private data to participate in FL training in order to improve the global FL model accuracy. Eventually, a driver can also maximize their own utility through contribution to either the training process or charging another EV having low level SoC in wireless way.

VIII. ACKNOWLEDGEMENTS

This work was supported by the Spanish Government under research project "Enhancing Communication Protocols with Machine Learning while Protecting Sensitive Data (COMPROMISE)" PID2020-113795RB-C31/AEI/10.13039/501100011033.

REFERENCES

[1] A. Ghosh, "Possibilities and challenges for the inclusion of the electric vehicle (ev) to reduce the carbon footprint in the transport sector: A review," *Energies*, vol. 13, no. 10, 2020. [Online]. Available: https://www.mdpi.com/1996-1073/13/10/2602

[2] *Global EV outlook 2019 - scaling-up the transition to electric mobility*, 2019.

[3] S. Zhang, W. Dou, Y. Zhang, W. Hao, Z. Chen, and Y. Liu, "A vehicle-environment cooperative control based velocity profile prediction method and case study in energy management of plug-in hybrid electric vehicles," *IEEE Access*, vol. 7, pp. 75 965–75 975, 2019.

[4] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2018.

[5] V. Mugunthan, A. Peraire-Bueno, and l. Kagal, "Privacyfl: A simulator for privacy-preserving and secure federated learning," in *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*, 2020.

[6] S. F. Lameh, W. Noble, Y. Amannejad, and A. Afshar, "Analysis of federated learning as a distributed solution for learning on edge devices," in *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, 2020, pp. 66–74.

[7] M. S. Sheikh, J. Liang, and W. Wang, "Security and privacy in vehicular ad hoc network and vehicle cloud computing: A survey," *Wireless Communications and Mobile Computing*, vol. 2020, no. 5129620, 2020.

[8] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: A viewpoint of vehicles as the infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, 2016.

[9] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular internet of things: Recent advances and open issues," *IEEE Open Journal of the Computer Society*, vol. 1, pp. 45–61, 2020.

[10] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Connected roads of the future: Use cases, requirements, and design considerations for vehicle-to-everything communications," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 110–123, 2018.

[11] M. Adil, J. Ali, Q. T. H. Ta, M. Attique, and T.-S. Chung, "A reliable sensor network infrastructure for electric vehicles to enable dynamic wireless charging based on machine learning technique," *IEEE Access*, vol. 8, pp. 187 933–187 947, 2020.

[12] Leticia Lemus Cárdenas, Ahmad M. Mezher, Pablo A. Barbecho Bautista, Mónica Aguilar Igartua, "A probability-based multimetric routing protocol for vehicular ad hoc networks in urban scenarios," *IEEE Access*, vol. 7, pp. 178 020–178 032, 2019.

[13] J. Posner, L. Tseng, M. Aloqaily, and Y. Jararweh, "Federated learning in vehicular networks: Opportunities and solutions," *IEEE Network*, vol. 35, no. 2, pp. 152–159, 2021.

[14] X. Mou, R. Zhao, and D. T. Gladwin, "Vehicle to vehicle charging (V2V) bases on wireless power transfer technology," in *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society*, 2018, pp. 4862–4867.

[15] "TensorFlow," https://www.tensorflow.org/federated/federated_learning.

[16] "OMNeT++," https://omnetpp.org/intro, accessed: 2020-10-06.

[17] "SUMO," https://www.eclipse.org/sumo/.

[18] "Veins," http://veins.car2x.org/.

[19] A. Hegde and A. Festag, "Artery-C: An OMNeT++ based discrete event simulation framework for cellular V2X," in *MSWiM '20: 23rd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2020.

[20] A. Virdis, G. Nardini, and G. Stea, "Modeling unicast device-to-device communications with simuLTE," in *2016 1st International Workshop on Link- and System Level Simulations (IWSLS)*, 2016, pp. 1–6.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Towards Flexible Integration of 5G and IIoT Technologies in Industry 4.0

Jorge Sasiain, Ane Sanz, Jasone Astorga, Eduardo Jacob
Department of Communications Engineering,
University of the Basque Country (UPV/EHU)
48013 Bilbao, Spain.
jorge.sasiain@ehu.eus, ane.sanz@ehu.eus, jasone.astorga@ehu.eus, eduardo.jacob@ehu.eus

Abstract
In this article, the authors propose a flexible network architecture for Industry 4.0 leveraging two 5G key-enabling technologies—Network Functions Virtualization and Software-Defined Networking. The authors also present the deployment of a Wireless Sensor Network with strong access control mechanisms into such architecture, enabling secure and flexible Industrial Internet of Things applications.

***Keywords*—industry 4.0, NFV, SDN, IIoT, 5G**

I. INTRODUCTION

Traditional network architectures present several shortcomings in regards to flexibility, accessibility, and dynamicity that can be further aggravated in industrial scenarios where the coexistence with heterogeneous devices, machine tools, and industry-specific protocols is a reality. The Industry 4.0 paradigm also envisions the integration of Industrial Internet of Things (IIoT) applications, often comprised of interconnected devices with limited capabilities and resources, named Constrained Device Sensors (CDS), which form a Wireless Sensor Network (WSN). However, the integration of WSNs into industrial environments comes with security issues that must be overcome.

The objectives of the present article are twofold. The first objective is to develop an NFV- and SDN-enabled architecture capable of overcoming the aforementioned limitations. The second objective is to integrate a WSN strengthened by the use of a lightweight, dynamic, and fine-grained access control protocol.

II. PROPOSED ARCHITECTURE

This section presents the deployment of the NFV- and SDN-enabled architecture for Industry 4.0, and the inclusion of a WSN for IIoT applications. It has been deployed across the Faculty of Engineering in Bilbao (EIB) and the Aeronautics Advanced Manufacturing Center (CFAA). These two locations are interconnected through a layer-2 SDN network at a rate of 10 Gbps. Two OpenStack nodes manage the resources of their respective locations, and three ONOS controllers manage the data plane connectivity. Finally, an Open Source MANO (OSM) on top orchestrates the whole NFV and SDN infrastructure. This architecture is represented in Figure 1.

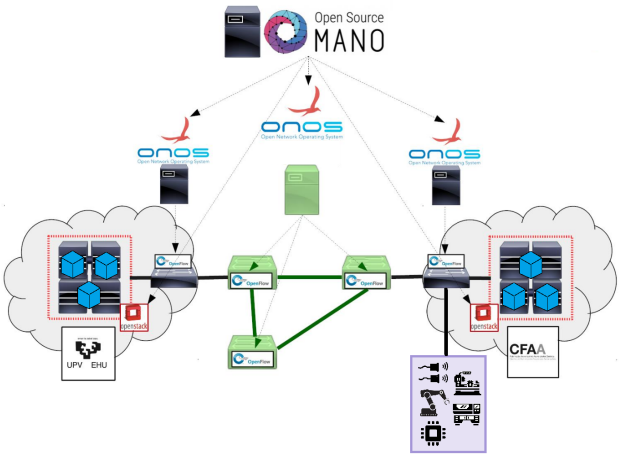


Fig. 1. Proposed NFV and SDN architecture

In addition, a WSN has been integrated into this architecture, enabling the deployment of several constrained sensors in order to monitor different environmental parameters. The Hidra protocol [1] is used, providing a strong and dynamic access control solution that can be implemented even in the most constrained devices.

Each Network Service (NS) deployed through OSM maps logically to a given manufacturing process and provides it, on demand, with the necessary resources and Virtual Network Functions (VNF). Layer-2 VLAN segmentation is used in order to ensure isolation between different processes. Besides, the addition of the WSN introduces an additional dimension to the design of NSs, as services provided by the CDSs can be dynamically allocated to these NSs together with the services provided by the VNFs that they are composed of.

ACKNOWLEDGEMENTS

This work was supported by 5G-City (TEC2016-76795-C6-5-R) and TRUE5G (PID2019-108713RB-C54).

REFERENCES

[1] Uriarte, M., Astorga, J., Jacob, E., Huarte, M., Carnerero, M. (2017). Expressive policy-based access control for resource-constrained devices. IEEE Access, 6, 15-46.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Mejorando la calidad de servicio en SDN mediante el ajuste dinámico del idle timeout con Deep Reinforcement Learning

Manuel Jiménez-Lázaro, Javier Berrocal, Jaime Galán-Jiménez
Escuela Politécnica de Cáceres,
Universidad de Extremadura
Avda. de la Universidad, S/N, Cáceres, España
{manueljimenez, jberolm, jaime}@unex.es

Las memorias TCAM *(Ternary content-addressable memory)* de las tablas de flujo de los nodos SDN *(Software Defined Networking)* son muy rápidas y permiten realizar búsquedas en paralelo en muy poco tiempo. Sin embargo, presentan un alto consumo de energía y un elevado coste, lo que hace que su tamaño sea limitado. Esta limitación de tamaño impacta sobre el número de reglas que se pueden instalar, por lo que una gestión ineficiente de las mismas puede suponer una degradación de la QoS *(Quality of Service)* de la red. Este trabajo propone una solución basada en DRL *(Deep Reinforcement Learning)* que permite ajustar dinámicamente el *idle timeout* de las reglas de flujo para maximizar el número de flujos que pueden ser encaminados en la red, lo que deriva en una mejora de la QoS.

***Palabras Clave*—SDN, DRL, idle timeout, TCAM.**

I. INTRODUCCIÓN

El nuevo paradigma de red SDN (*Software Defined Networking*) separa el plano de datos del plano de control, estando gestionado por un elemento centralizado que tiene una visión global de la red. Este elemento centralizado, denominado controlador, es quien se encargará, entre otras cosas, de establecer el encaminamiento entre nodos. Eso lo hará guardando las reglas de encaminamiento en las tablas de todos los nodos del flujo. Cuando llega un paquete a un *switch* SDN, tiene que encaminarse en función de las reglas almacenadas en la tabla de flujo de ese nodo. Si el paquete (con sus diferentes campos de las distintas cabeceras) hace *match* con una de las reglas de la tabla, podrá enviarse al siguiente nodo en el camino, acercándose a su destino. Si no existe regla relacionada, el paquete no podrá ser encaminado inmediatamente (lo que se conoce como *table-miss*), y el *switch*, tendrá que preguntar al controlador qué hacer con dicho paquete (mensaje PACKET_IN) para instalar el flujo. El controlador le indicará qué reglas debe instalar y en qué *switches* de forma que permita a los datos llegar a su destino. Esto, lógicamente, supone una penalización de tiempo, al aumentar el RTT (*Round-trip time*) del primer paquete del flujo. Además, debido a que el número de reglas que se pueden instalar en los *switches* SDN es limitado [1], [2], se puede dar la situación en la que no existe regla que haga *match* con el paquete pero tampoco hay espacio disponible en la tabla de flujo para poder instalarla, por lo que este flujo no podrá ser encaminado y se perderán los datos enviados. Por lo tanto, se traduciría en una reducción de la QoS (*Quality of Service*) de la red.

Adicionalmente, es importante comentar que el *match* que hace un paquete entrante con una regla en un entorno SDN no solo tiene en cuenta los campos dirección IP origen y dirección IP destino como se muestra en este trabajo. SDN es capaz de tener en cuenta, además, varios campos de las diferentes cabeceras (puertos, protocolo, etc). Las tablas de flujo se implementan mediante la utilización de memorias TCAM. Estas memorias TCAM son muy rápidas, pues utilizan un sistema de búsqueda paralela. Esto es algo muy importante cuando queremos encaminar paquetes lo antes posible. Sin embargo, hacen uso de mucha energía [3], tienen un coste muy alto [4] y, como consecuencia, disponen de un espacio bastante limitado [5], [6], algo que repercutirá negativamente a la hora de encaminar los paquetes, pudiendo almacenar menos reglas al mismo tiempo.

Además, existen los conceptos de *idle timeout* y *hard timeout*. Para evitar que una regla se quede almacenada en una TCAM durante demasiado tiempo sin ser utilizada, cada regla incluye estos dos valores. Si la regla se mantiene en la tabla sin ser usada por una duración igual al *idle timeout*, la regla será borrada ya que no se considerará lo suficiente activa, dejando espacio a otras posibles reglas. El valor del *idle timeout* normalmente se establece de forma estática para cada regla. Es decir, una

vez que se añade ese valor, se mantendrá así durante toda la estancia de la regla en la TCAM. El *hard timeout* es similar, solo que borrará la regla cuando iguala la duración del *hard timeout*, empezando a contar desde el momento en el que se instaló la regla y sin reiniciarse al ser utilizada como en el caso del *idle timeout*.

En los últimos años, existe un interés creciente en aprovechar las capacidades de las redes SDN para mejorar la QoS ofrecida. Trabajos como los de [2], [7] proponen soluciones para mejorar la eficiencia energética de las redes SDN considerando las restricciones impuestas por el tamaño limitado de las memorias TCAM de las tablas de flujo. Esta limitación también impacta sobre el rendimiento de las redes SDN/NFV en las que se utiliza el paradigma SFC (*Service Function Chaining*), ya que los nodos que actúan como clasificadores pueden verse restringidos en el número de solicitudes SFC que pueden aceptar. Para ello, trabajos como los de [6], [8], [9] proponen realizar una descarga en el proceso de clasificación. Dicha acción podrá ser realizada por cualquiera de los nodos del camino desde el nodo origen (clasificador inicial) hasta el nodo destino.

En este trabajo, nuestro objetivo es ajustar dinámicamente el *idle timeout* para mejorar la QoS de las redes SDN. Si instalamos en cada regla *idle timeouts* altos, estamos provocando que haya menos intercambio de información con el controlador por el simple hecho de que las reglas no se borrarán a menudo, lo que implica que no tendrán que reinstalarse demasiadas veces. Pero esto supondrá que las TCAM se llenarán pronto, por causa de que las reglas tardarán en borrarse, no pudiendo satisfacer nuevos flujos.

Si hacemos lo contrario e instalamos *idle timeouts* bajos, el intercambio de mensajes de control de tipo *OpenFlow* con el controlador aumentará enormemente, suponiendo retrasos en la red, mas podremos satisfacer un mayor número de flujos porque no habrá reglas que acaparen las TCAM durante un gran espacio de tiempo.

El objetivo de este trabajo es encontrar una manera eficiente de tratar los *idle timeouts* de cada regla de la red de forma dinámica. Así, queremos optimizar los valores que se le adjudican a cada *idle timeout* para conseguir un mejor funcionamiento global de nuestra red, estableciendo una mejor administración del espacio de las TCAM, y por tanto, maximizando la cantidad de flujos que se pueden satisfacer. Para ello, se propone un algoritmo basado en DRL que recorrerá las reglas instaladas e irá ajustado sus *idle timeouts* en función del estado de la red, para así conseguir que estos *idle* tengan valores acordes a las necesidades de la red.

Los resultados obtenidos tras la ejecución de simulaciones sobre redes sintéticas y reales indican que nuestro algoritmo mejora, por norma general, el desempeño de la red en cuanto a la cantidad de paquetes que pueden satisfacerse, consiguiendo una mejor QoS.

El resto del artículo se organiza de la siguiente manera: el algoritmo DRL propuesto se describe en la Sección II, la Sección III describe los resultados obtenidos tras la simulaciones realizadas y las comparaciones con la forma estática tradicional de usar los *idle timeout*, y por último, la Sección IV hace un repaso de las conclusiones obtenidas.

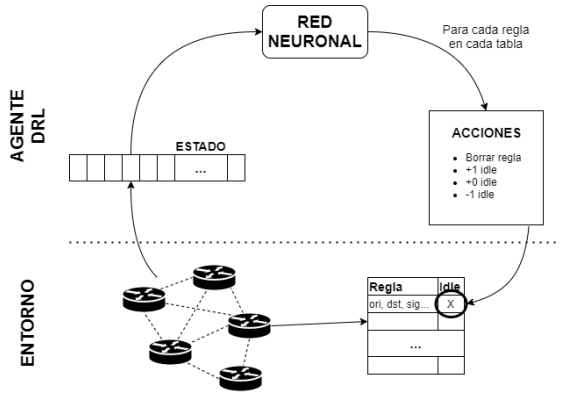


Fig. 1: Esquema del funcionamiento del algoritmo DRL

II. ALGORITMO DIDLE DRL

Reinforcement Learning es un subcampo de ML (*Machine Learning*), que se diferencia de los métodos más clásicos como los supervisados [10] y los no supervisados [11]. Los supervisados se centran en clasificar un conjunto de datos de entrada, y los no supervisados en agrupar ese conjunto de entrada. Sin embargo, *Reinforcement Learning* [12] se basa en entrenar a un agente para que tome ciertas acciones sobre un entorno concreto en función de las recompensas que se le devuelven, pudiendo encontrar soluciones a problemas concretos.

DRL es parte de *Reinforcement Learning*, pero incluye la palabra *Deep*, pues hace uso de una red neuronal, cuyo objetivo último será aprender cómo maximizar las recompensas que obtiene [13].

DRL está compuesto por varios conceptos. Uno de ellos es el *entorno*, que se refiere a aquello que el algoritmo utiliza en su proceso de predicción, sobre el que realiza las *acciones* y sobre el que obtiene las *recompensas*. En nuestro caso, el entorno es la configuración de la topología de la red con el estado y contenido de las tablas de flujo de los nodos SDN.

El *agente* es el ente que toma las decisiones (la red neuronal) y que pretende aprender para optimizar dichas decisiones de forma que se maximice la *recompensa* que se le devuelve tras la toma de una *acción*. Por ello, el siguiente concepto es la *acción*, que es cada modificación que el agente realiza sobre el *entorno*. La recompensa que se le ofrece a la red neuronal tras cada *acción* depende de una evaluación que se hace sobre el *entorno*, identificando en este si la situación en la que se encuentra tras la acción es positiva o negativa, lo que devolverá una recompensa acorde a dicha situación. En nuestro algoritmo, una situación positiva será aquella en la que se pueden cumplir la mayor cantidad de flujos posibles de cara al siguiente instante de tiempo, a la vez que las TCAM se encuentran con una ocupación baja. Si las TCAM están llenas, y no se pueden satisfacer las siguientes reglas, la recompensa será negativa.

A continuación, se procede a explicar con mayor profundidad el funcionamiento particular de nuestro algoritmo *DIdle DRL (Dynamic Idle DRL)*.

A. Espacio de estados

El estado es el conjunto de datos representativos de la situación actual de la red que la red neuronal recibe como entrada, y el cual usará para tomar unas acciones u otras. En la Figura 1 podemos ver representado el estado por un vector, el cual está compuesto por:

- MT: La matriz de tráfico del instante siguiente. Se trata de una matriz conformada por unos y ceros, indicando un 1 que hay actividad entre el nodo origen, representado por el número de la fila, y el nodo destino, representado por el número de la columna. Un 0 indica que no hay actividad, que no se quiere intercambiar información entre ambos nodos.
- $\mathcal{R}$: El conjunto de todas las reglas instaladas en todos los nodos. Cada regla con su nodo origen, su nodo destino, su siguiente salto, sus contadores y su *idle timeout*. $r = [src\_node, dst\_node, next\_hop, counters, idle\_timeout]$

Este espacio de estados lo recibirá la red neuronal, que como salida devolverá una acción, como se observa en la Figura 1, en función que lo que haya aprendido a raíz de las recompensas otorgadas durante su fase de entrenamiento.

B. Acciones

A la hora de instalar una regla en una tabla de flujo, se incluirá con un determinado *idle timeout* y DRL no intervendrá en ello, porque el algoritmo solo modifica los *idle timeout* de reglas ya instaladas. Una vez finaliza un instante de tiempo, habiéndose incluido las nuevas reglas que ese instante de tiempo requería y habiéndose realizado el envío de los datos que corresponden, se hace una llamada al *Algoritmo DIdle DRL*. El pseudocódigo de *DIdle DRL* viene descrito en Alg. 1, donde se observa que se le pasa como entrada el estado actual de la red y se le devuelve una de las cuatro acciones posibles. Las acciones se llevarán a cabo para cada regla de cada TCAM, es decir, el algoritmo tendrá que recorrer todas y cada una de las reglas que están instaladas en nuestra red y actualizar sus respectivos *idle*. Como resultado, se obtiene un nuevo estado de la red.

Las distintas acciones que puede tomar para cada regla son:

- *Borrar la regla directamente*. Si DRL considera que una determinada regla no está siendo útil y está ocupando espacio, puede borrarla.
- *Disminuir en una unidad el idle timeout de la regla.*
- *Mantener el idle timeout de la regla.*
- *Aumentar en una unidad el idle timeout de la regla.*

C. Recompensas

Las recompensas con las que se retroalimenta DRL son esenciales para un buen funcionamiento del algoritmo. En nuestro caso se ha tomado la decisión de establecer una recompensa que evalúe el estado de la red una vez se han llevado a cabo las acciones sobre todas las reglas. El objetivo último de DRL será maximizar el valor de esa recompensa, por lo que una recompensa alta reforzará los pesos de la red neuronal para favorecer ese resultado más a menudo.

Algorithm 1 Pseudocódigo del ajuste de los *idle timeouts*.

Require: Un vector que representa el estado de la red, formado por la matriz de tráfico del siguiente instante: MT, y por el conjunto de reglas instaladas en cada TCAM: $\mathcal{R}$.

1: **for all** $r \in \mathcal{R}$ **do**
2: Tomar acción A ▷ La acción la toma en función del estado recibido
3: Actualizar r con la acción A
4: **end for**
5: **return** *Estado con las reglas actualizadas por las acciones*

Este valor de recompensa (rec) se calcula de acuerdo a la ecuación 1:

$$\text{rec} = Tam_{TCAM} - Max_{TCAM} - \sum Activo_{NO} \quad (1)$$

Por un lado, se calcula el máximo de ocupación que tiene una TCAM tras realizar las acciones (Max_{TCAM} en la ecuación 1). Al tamaño que tiene una TCAM (Tam_{TCAM} en la ecuación 1) de la red se le resta ese valor calculado, y esa será la recompensa que tendremos hasta el momento. Con esto conseguimos minimizar la ocupación de las TCAM para tratar de que exista espacio suficiente para nuevas reglas.

Por otro lado, se analiza la matriz de tráfico del instante que va a comenzar a continuación, y se hace un sumatorio de todos los flujos que tienen que ser insertados como reglas en las distintas TCAM como consecuencia de que en el instante actual no se encuentran instalados ($Activo_{NO}$ en la ecuación 1). Esto se lo restamos a la recompensa que teníamos del punto anterior. Restar este número ayuda a DRL a no borrar reglas que van a ser utilizadas a continuación, porque tener que instalarlas nuevamente produce una pérdida de tiempo por los intercambios de información con el controlador SDN, lo que repercute negativamente en el funcionamiento de la red.

Así, equilibraríamos el hecho de querer minimizar la ocupación de las tablas con el hecho de satisfacer la mayor cantidad de flujos sin que se tengan que reinstalar nuevamente.

D. Cálculo del idle dinámico

Con el objetivo de realizar un tratamiento completamente dinámico de los *idle timeout*, proponemos una fórmula para establecerlo a la hora de insertar las reglas en las TCAM, teniendo en cuenta el uso histórico del flujo y la ocupación de las TCAM.

$$\text{idle timeout} = (\frac{P_x}{100} + k) * \frac{N_{dis}}{N_{tot}} * \frac{t_{max}}{2} \quad (2)$$

A continuación se explican cada uno de los términos de la ecuación 2:

- P_x: Porcentaje de uso del flujo de la regla a instalar en los últimos x instantes de tiempo.
- k: Valor constante que se encuentra entre -1 y 1. Sirve para ajustar el *idle timeout* en función de cómo queremos que se comporte la fórmula. Si el valor es positivo, los valores de los *idle timeout* serán mayores. Si lo situamos negativo, serán más pequeños.
- N_{dis}: Número de espacios disponibles en la tabla de encaminamiento más ocupada del flujo a insertar.
- N_{tot}: Número de espacios totales que posee una tabla de encaminamiento (lo consideramos el mismo para todas las tablas de la red).
- t_{max}: Número de instantes de tiempo totales en la simulación. El mínimo *idle timeout* que se puede asociar a una regla es de 1, y el máximo consideraremos que es la mitad de todos los instantes de tiempo que estamos recorriendo en la simulación. De forma que el *idle timeout* resultante debe estar entre 0 y el instante de tiempo máximo dividido entre dos.

Para un correcto funcionamiento del algoritmo, para el valor resultante de la ecuación 2 debe aplicarse la función techo. Es decir, si obtenemos un *idle timeout* de 2,1, debemos redondearlo a 3. Por supuesto, si la k es positiva, puede darse el caso de que el *idle timeout* resultante sea superior al valor máximo que hemos puesto como límite $(t_{max}/2)$. En ese caso, podemos limitar que todo valor que supere el límite se reduzca a este, o podemos dejar que el *timeout* lo supere, dependerá de nuestros intereses. Si la k es negativa, se producirá el efecto contrario. En caso de que se obtenga un valor negativo, el *idle timeout* se establecerá a 0, lo que significa que se usará la regla e inmediatamente se borrará.

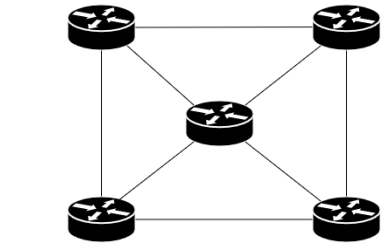
III. RESULTADOS EXPERIMENTALES

Para las fases de entrenamiento de nuestro *algoritmo DIdle DRL*, se han generado los estados que recibe la red de una forma pseudoaleatoria, de forma que representen estados reales que se podría encontrar a la hora de enfrentarse al problema real. Para este entrenamiento se han utilizado 50000 pasos.

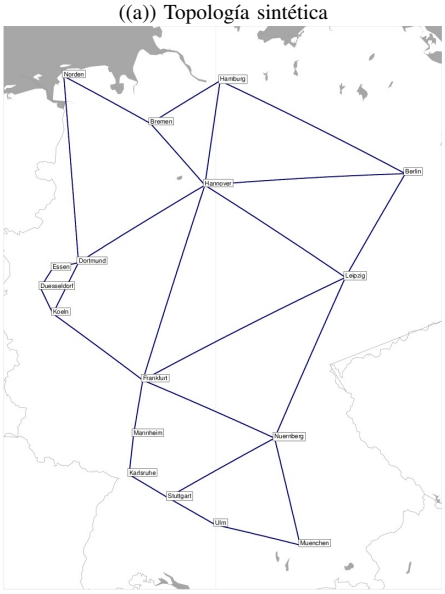
Tras las fases de entrenamiento del algoritmo DRL, se ejecutaron una serie de pruebas que nos indicaron el funcionamiento de este con respecto al funcionamiento estándar. El funcionamiento estándar es el que tendría la red en caso de que los *idle timeout* se mantuvieran de manera estática durante toda la estancia de la regla en la TCAM.

Las pruebas se realizaron sobre dos topologías de red diferentes. Una topología sintética de 5 nodos y 8 enlaces bidireccionales que puede verse en la Figura 2(a) y una topología real obtenida de la librería SNDLib, denominada Nobel-Germany (17 nodos y 25 enlaces) representada en la Figura 2(b).

Con respecto al tráfico generado, consideramos dinamicidad en el mismo, intercalando intervalos de actividad con intervalos de inactividad. Para ello nos vamos en DTMP (*Discrete Time Markov Process*), que generará la actividad



((a)) Topología sintética



((b)) Nobel-Germany

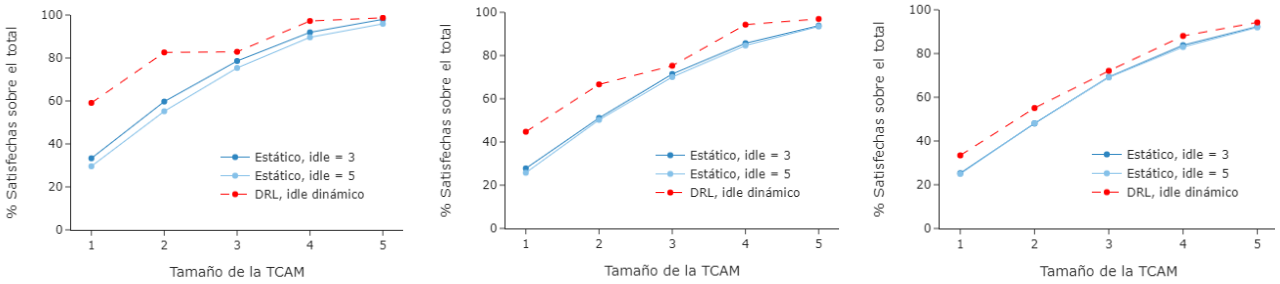
Fig. 2: Topologías de red consideradas.

en función de las probabilidades que otorgan los valores de α y β, siendo $\beta = 1 - \alpha$. Para cada topología haremos tres pruebas diferentes, una para una carga de tráfico del 25%, que correspondería con un $\alpha = 0.25$, otra para el 50%, siendo $\alpha = 0.5$ y otra para el 75%, con $\alpha = 0.75$.

Volviendo a las pruebas, la simulación que realizamos para visualizar si el algoritmo DRL es exitoso ha sido programada a través del lenguaje *Python*, donde generamos la matriz de tráfico mediante DTMP para un número X de instantes de tiempo. En nuestro caso, las pruebas las hemos realizado para 10 instantes de tiempo, para un valor de $k = 0.1$ y $x = 3$ en la fórmula que calcula el *idle timeout* inicial de una regla.

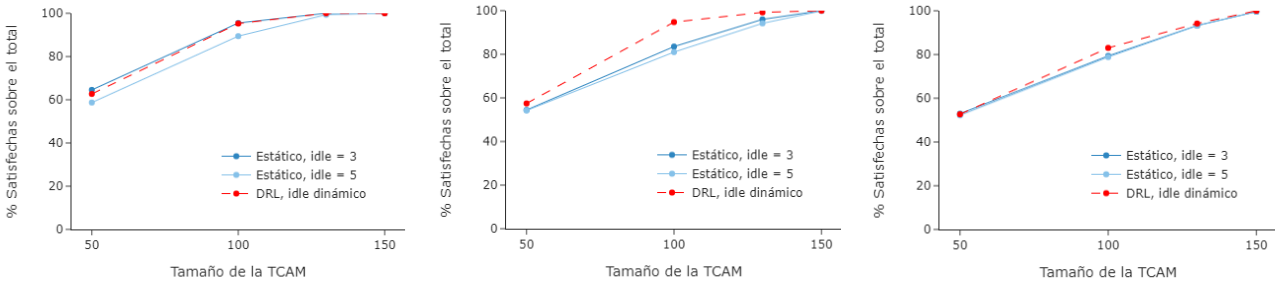
En cada instante se insertarán todas las reglas de los flujos activos según la matriz de tráfico, a excepción de aquellos flujos en los que alguna de las TCAM de los nodos por los que pasa se encuentre completamente ocupada. Una vez insertadas todas las reglas posibles, se envía la información y se llama a *DIdle DRL* para que lleve a cabo las acciones que considere convenientes, ajustando los *idle timeouts* o borrando reglas. Tras esto, se eliminan las reglas cuyo tiempo de inactividad haya llegado al *idle timeout* y se repite el bucle.

Para obtener los resultados de cada gráfica, haremos varias iteraciones de la simulación para así poder hacer una media entre todas y evitar casos extraños o extremos. En el caso de la topología sintética de la Figura 2(a) hemos



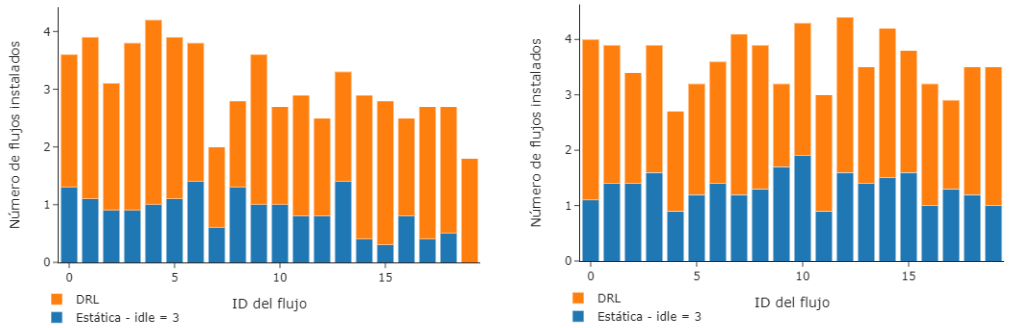
((a)) $\alpha = 0.25$ ((b)) $\alpha = 0.5$ ((c)) $\alpha = 0.75$

Fig. 3: Porcentaje de flujos instalados correctamente en la topología sintética.



((a)) $\alpha = 0.25$ ((b)) $\alpha = 0.5$ ((c)) $\alpha = 0.75$

Fig. 4: Porcentaje de flujos instalados correctamente en Nobel.



((a)) Tamaño de TCAM = 2 ((b)) Tamaño de TCAM = 4

Fig. 5: Comparación del número de veces que ha de instalarse un flujo en DRL y en estático para $\alpha = 0.25$ en la topología sintética.

hecho 100 iteraciones y hemos calculado la media de sus resultados. En la topología nobel de la Figura 2(b) han sido 10 iteraciones por su mayor complejidad.

Las primeras gráficas que nos interesan conocer son las que podemos ver en la Figura 3 para la topología sintética y en la Figura 4 para la topología Nobel. Podemos ver como estas seis gráficas muestran el resultado para el uso de DRL con el *idle* dinámico, para un *idle* estático igual a 3 y para un *idle* estático igual a 5. *Idle* estático es la forma tradicional de tratar con estos valores, iniciando una regla con un *idle timeout* determinado, sin posibilidad de cambio hasta que las reglas de ese flujo se eliminen de sus TCAM.

Hemos establecido esos valores estáticos entendiendo que un *idle timeout* de 3 es lo mínimo aceptable, ya que un *idle* de 1 sabemos que funcionará muy bien por el hecho de que siempre estará borrando reglas y habrá espacio para nuevas, por lo que es muy difícil que pueda ser superado por DRL en cuanto a flujos satisfechos. Por otra parte, tampoco tiene sentido evaluar para un *idle* de 1 porque es algo que a nivel práctico no se utiliza al ser un valor muy pequeño.

Estas gráficas nos indican el porcentaje de flujos que se

satisfacen del total de flujos que quieren enviarse, entendiendo como flujo el conjunto de reglas que permiten a un paquete llegar desde un nodo origen a su nodo destino. Para verlo de una manera más explicativa, imaginemos que el nodo 1 quiere enviar información al nodo 3, y el nodo 4 quiere enviar información al nodo 2. Si los dos flujos se satisfacen pues hay espacio en las TCAM correspondientes, tendríamos un porcentaje del 100% de satisfechas. Si uno de los dos no pudiera cumplirse, sería del 50%.

Comprobamos que en las gráficas de la Figura 3 de la topología sintética, DRL funciona mejor que de la forma estática en todos los casos, aunque esta diferencia se reduce cuanto más grande es el tamaño de la TCAM. Podemos deducir entonces que nuestro algoritmo funciona mejor cuando se enfrenta a tamaños de TCAM inferiores, lo que es un resultado positivo pues nuestro problema se basa en TCAM de tamaños limitados. Añadiendo a lo anterior, cuanto más grande es la TCAM vemos que los *idle* estáticos se acercan al 100%, satisfaciendo todos los flujos en la red.

Como ampliación de lo anterior, también se encuentran diferencias entre las gráficas con un valor de α inferior, como la Figura 3(a), con respecto a las que tienen mayor carga de tráfico, como la Figura 3(c). La diferencia reside en que cuanta menor carga, más libertad tiene el algoritmo de jugar con las reglas instaladas, pudiendo aumentar con más éxito el número de flujos satisfechos. Esto se complica en una prueba con más carga de tráfico, pues es muy posible que la TCAM se encuentre repleta de reglas que van a utilizarse, a la vez que hay paquetes a enviar que necesitan reglas que no tienen espacio disponible. Entonces, se complica la capacidad que tiene el algoritmo de jugar con esas reglas, reduciendo la diferencia ante la simulación estática.

Para la figura 4 con las gráficas de Nobel, vemos una mejora menor, siendo los resultados más ajustados que en la topología sintética, llegando a estar igualados en situaciones de baja actividad de tráfico (Figura 4(a)).

Viendo que el resultado general de las gráficas que comparan el grado de flujos satisfechos es positivo, se va a observar mediante la Figura 5 cómo repercute en el número de flujos instalados el hecho de que el número de flujos satisfechos sea mayor.

En estas gráficas se va a analizar cuántas veces se instala cada regla de media, estando representada cada regla por un id. El resultado ideal sería que este número de veces que se instala cada regla fuera igual o menor respecto a los *idle* estáticos, pero vemos en las dos gráficas que no es así, que las veces que hay que instalar cada regla es, por norma general, mayor en *DIdle DRL*. Aún así, se observa que, al relajar el tamaño de la TCAM (Figura 5(b)) con respecto a un tamaño menor (Figura 5(b)), el número de instalaciones baja en DRL, acercándose al número de instalaciones del tratamiento estático del *idle*.

En este aspecto, no cumplimos con la imagen ideal que teníamos, pero se puede entender que es una penalización que tenemos que pagar, porque el objetivo principal es poder instalar las reglas en las tablas de flujos para que el tráfico se pueda encaminar, aunque tengamos que realizar un mayor número de comunicaciones con el controlador.

IV. CONCLUSIONES

El objetivo de este trabajo era encontrar una manera eficiente de tratar los *idle timeouts* de cada regla de la red de una forma dinámica, estableciendo una mejor administración del espacio de las TCAM, y entonces, maximizando la cantidad de flujos que se pueden satisfacer. Con esto en mente, se propuso un algoritmo DRL, denominado *DIdle DRL*, que recorría las reglas instaladas e iría ajustando los *idle timeouts* en función del estado de la red en ese momento.

Tras las pertinentes pruebas, comprobamos que los resultados obtenidos eran positivos, y que *DIdle DRL* mejoraba normalmente el número de paquetes que podían llegar a su destino.

En contraposición, se encontraba un aumento de reglas a instalarse. Ello suponía un aumento de intercambio de mensajes con el controlador, pero se ha entendido como un mal menor a cambio de conseguir que la información consiga llegar al destino en unos espacios de TCAM tan limitados.

En definitiva, se ha logrado el objetivo de mejorar la QoS de la red mediante el algoritmo *DIdle DRL*.

AGRADECIMIENTOS

Este trabajo ha sido financiado, en parte, por el proyecto RTI2018-094591-B-I00 (MCI/AEI/FEDER,UE), el proyecto 4IE+ (0499-4IE-PLUS-4-E) financiado por el programa Interreg V-A Espana-Portugal (POCTEP) 2014-2020, por la Consejería de Economía, Ciencia y Agenda Digital de la Junta de Extremadura (GR18112, IB18030) y por el Fondo Europeo de Desarrollo Regional (FEDER).

REFERENCIAS

[1] A. R. Curtis, J. C. Mogul, J. Tourrilhes, P. Yalagandula, P. Sharma, and S. Banerjee, "Devoflow: Scaling flow management for high-performance networks," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 4, pp. 254–265, aug 2011.

[2] J. Galán-Jiménez, J. Berrocal, J. L. Herrera, and M. Polverini, "Multi-objective genetic algorithm for the joint optimization of energy efficiency and rule reduction in software-defined networks," in *2020 11th International Conference on Network of the Future (NoF)*, 2020, pp. 33–37.

[3] C. R. Meiners, A. X. Liu, and E. Torng, "Bit weaving: A non-prefix approach to compressing packet classifiers in tcams," *IEEE/ACM Transactions on Networking*, vol. 20, no. 2, pp. 488–500, April 2012.

[4] P. C. Lekkas, *Network Processors: Architectures, Protocols, and Platforms*. New York, NY, USA: McGraw-Hill, 2003.

[5] M. Kuźniar, P. Perešíni, D. Kostić, and M. Canini, "Methodology, measurement and analysis of flow table update characteristics in hardware openflow switches," *Computer Networks*, vol. 136, pp. 22–36, 2018.

[6] M. Polverini, J. Galán-Jiménez, F. G. Lavacca, A. Cianfrani, and V. Eramo, "A scalable and offloading-based traffic classification solution in nfv/sdn network architectures," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1445–1460, 2021.

[7] J. Galán-Jiménez, J. Berrocal, M. Linaje, and C. Gómez, "Minimización del consumo de energía en redes sdn bajo restricciones tcam," in *Actas de las XIV Jornadas de Ingeniería Telemática (JITEL 2019)*, 2019, pp. 1–6.

[8] M. Polverini, J. Galán-Jiménez, F. G. Lavacca, A. Cianfrani, and V. Eramo, "Dynamic in-network classification for service function chaining ready sdn networks," in *2019 10th International Conference on Networks of the Future (NoF)*, 2019, pp. 74–81.
[9] ——, "Improving dynamic service function chaining classification in nfv/sdn networks through the offloading concept," *Computer Networks*, vol. 182, p. 107480, 2020.
[10] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging artificial intelligence applications in computer engineering*, vol. 160, no. 1, pp. 3–24, 2007.
[11] A. Kassambara, *Practical guide to cluster analysis in R: Unsupervised machine learning*. Sthda, 2017, vol. 1.
[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
[13] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Towards integrating hardware Data Plane acceleration in Network Functions Virtualization

David Franco[1], Asier Atutxa[1], Jorge Sasiain[1], Eder Ollora[2], Marivi Higuero[1], Jasone Astorga[1], Eduardo Jacob[1]
[1]Department of Communications Engineering, University of the Basque Country (UPV/EHU). 48013 Bilbao, Spain.
{david.franco, asier.atutxa, jorge.sasiain, marivi.higuero, jasone.astorga, eduardo.jacob}@ehu.eus
[2]DTU Fotonik, Technical University of Denmark. Kongens Lyngby, Denmark.
eoza@fotonik.dtu.dk

***Abstract*—This paper proposes a framework for integrating data plane (DP) acceleration within the Network Functions Virtualization (NFV) architecture. Data plane programming (DPP) proves to be beneficial for NFV environments, as it provides full packet forwarding flexibility through the use of self-designed algorithms. Additionally, DPP provides high-performance networking, as the DP can be configured to execute specific functions on dedicated hardware. We present an integration of the DP acceleration within the ETSI NFV architecture that leverages custom DP functions implemented in hardware switches using P4 language. Besides, OpenStack and Kubernetes are used as Virtualized Infrastructure Managers (VIMs) and Open Source MANO (OSM) as the Management and Orchestration (MANO) element.**

***Keywords*—P4, NFV, data plane acceleration**

I. INTRODUCTION

In the last years, Network Functions Virtualization (NFV) and Software-Defined Networking (SDN) have changed the framework for the deployment of services. On the one hand, NFV allows Network Functions (NFs) to be deployed as Virtual Network Functions (VNFs) over a commercial off-the-shelf hardware infrastructure. On the other hand, SDN solves the problem of having a vendor-specific control plane (CP) in network devices, allowing the definition of custom CPs designed by the network operator. However, full packet forwarding flexibility is given by data plane programming (DPP), which can forward packets following self-designed algorithms and perform custom actions to packets with user-defined formats. Programming Protocol-Independent Packet Processors (P4) is a DPP language that provides a high abstraction level to define packet processing pipelines. DPP also provides high performance networking, as the data plane of network devices can be configured to execute specific functions on dedicated hardware. The addition of DPP to NFV through hardware network devices enhances the performance in the communication among VNFs thanks to line-rate packet processing capability and the ability to offload certain compute-intensive functions.

This paper proposes an integration of the DP acceleration within the standard ETSI NFV architecture, using OpenStack and Kubernetes to manage the NFV Infrastructure (NFVI), and Open Source MANO (OSM) as the top-level NFV orchestrator. The framework integrates the lifecycle of P4-enabled hardware switches in the NFV architecture to enable the automatic offloading of NFs to the DP.

II. RELATED WORK

Different works have been presented in the literature regarding the data plane (DP) function offloading in NFV. For instance, [1], [2] introduce two frameworks to offload VNF processing to P4-based Network Interface Cards (NICs) and software switches respectively. The former decomposes VNFs into small embedded NFs that are offloaded to the DP, and the latter runs NFs as P4 programs on Docker containers that implement P4 targets.

Authors in [3] propose an ETSI NFV-based architecture that allows the instantiation of P4-based NFs. They explain the necessity of reconfiguring the DP when multiple users are sharing the same physical P4 target and they test their architecture over P4 software switches. In this sense, [4], [5], [6] provide different approaches to achieve DP modularity and allow a transparent reconfiguration of the DP. The majority of them require modifications of the P4 compiler, or even in the architecture of the P4 target. For instance, [5], [6] present abstraction layers and mechanisms to create modular P4 programs that can then be merged.

Authors in [7], [8] describe how to integrate P4 within OpenStack, by modifying the Neutron module to allow the offloading of some VNF functions to the DP of P4 software switches. They classify their VNF offloading according to the acceleration techniques defined by the ETSI NFV [9]. However, none of these approaches considers the use of hardware-based targets.

This paper focuses on the advantages of integrating P4-enabled hardware switches in NFV to automate the offloading and acceleration of custom NFs to the DP.

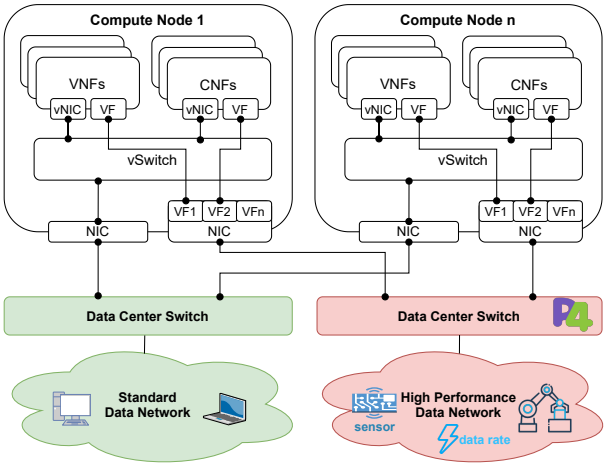


Fig. 1. Proposed system architecture to integrate P4NFs.

III. PROPOSED FRAMEWORK FOR DP ACCELERATION IN NFV

This section describes the proposed integration of the DP acceleration within the ETSI NFV architecture. The proposed framework allows the users to create network services (NSs) that leverage custom DP functions implemented in P4 language. For this purpose, physical P4-enabled switches are integrated into the NFVI.

A. Functional description

In this subsection we present an overview of the proposed technologies and approaches to enable the deployment of NSs combining traditional NFs —VNFs and Container Network Functions (CNFs)— with our novel proposed concept of P4 Network Functions (P4NFs). We propose the deployment of such P4NFs over P4 switches.

As part of the NFV architecture, OpenStack and Kubernetes are used as Virtualized Infrastructure Managers (VIMs) and Open Source MANO (OSM) is used as the Management and Orchestration (MANO) element. The NFVI is composed of several compute nodes that form the OpenStack cloud and the Kubernetes cluster, as well as of the switching infrastructure to interconnect those nodes, which includes both regular and P4 switches. OSM can deploy VNFs and CNFs on top of OpenStack and Kubernetes respectively, whereas P4NFs are employed to configure the P4 switches. These switches are registered in OSM as Physical Deployment Units. Connectivity between VNFs and CNFs can be provided through the regular top of the rack switches and/or through the P4 switches. To make this possible, the compute nodes have network interfaces connected to both types of switches, enabling the VNFs/CNFs to connect to any network. The VNF/CNF interfaces towards the P4 switches are SR-IOV Virtual Functions (VFs). To achieve isolation between tenants —i.e. between NSs of different tenants—, separate OpenStack VLAN networks are assigned to each tenant. This proposed architecture is depicted in Fig. 1.

To design a P4NF, the user provides a reference to the desired functions in a P4NF metadata file, such as simple

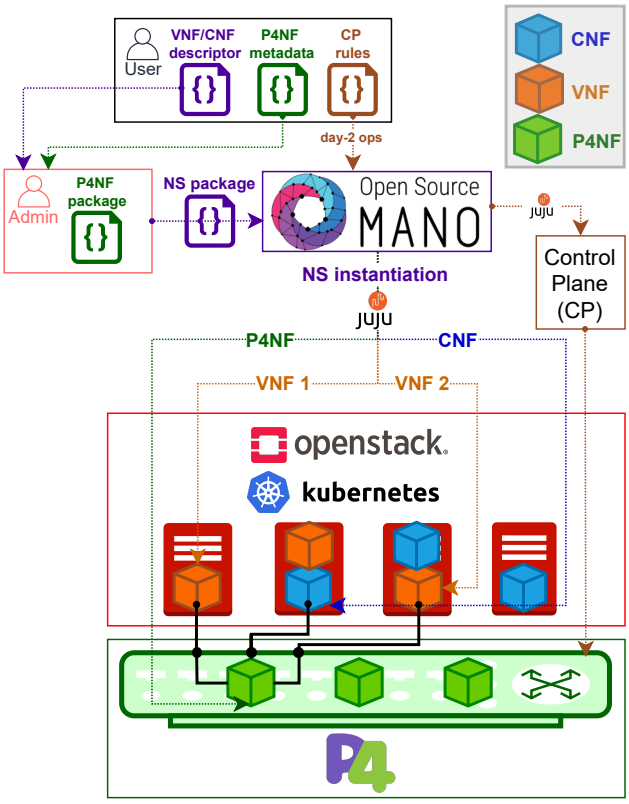


Fig. 2. P4-based NS onboarding and instantiation process.

layer 2 forwarding, or layer 3 routing plus a custom layer 3 firewall. These functions are extracted from a shared repository of P4 code in which they are implemented in a modular way. A user may upload a new function that is not yet stored in the repository to use it in the P4NF. The requested functions are contrasted with the current state of the P4 pipeline, and the switch is reconfigured in order to upgrade to a new state that integrates the requirements of the new P4NF, which involves recompiling the DP and reconfiguring the CP. Every P4NF must include the parsing of Ethernet and VLAN to achieve the aforementioned tenant-level isolation.

The deployment of these P4NFs leverages OSM's support of Juju charms, which encapsulate a set of Python scripts that can be executed towards the target of the P4NF, i.e. a P4 switch. Apart from the P4NF, the user encapsulates the desired VNFs/CNFs into an NS. The P4NF designer provides the network administrator with the P4NF metadata file plus the VNFs/CNFs, so that the latter builds the complete package and instantiates it. Furthermore, Juju charms can be used to implement day-2 operations that allow the P4NF designer to directly modify the CP rules at any time once the NS is created. An example of the onboarding, instantiation, and lifecycle management processes of a composite NS is illustrated in Fig. 2.

B. Low-level design

This subsection presents a low-level description of the proposed deployment methodology for DP acceleration

in NFV. Specifically, a workflow for the lifecycle management of the envisioned NSs is analyzed, detailing the procedures involved in each task.

The framework hereby presented provides three main features that enable the description, instantiation, and management of the P4-enabled NSs. It provides **(a) multi-tenancy**, so the connectivity between NFs is restricted to NSs connected to the same OpenStack VLAN network, and, thus, to NSs belonging to the same tenant. It also offers **(b) DP composition** to describe P4 functions at NS level and then combine them to create a single P4 program that comprises the P4 functions of all running NSs —even those from different tenants. In this way, the DP must be recomposed each time an NS is instantiated or terminated. The DP composing is implemented according to [5], [6]. Lastly, it supports **(c) DP slicing** to implement different and isolated packet processing pipelines for each NS that is configured when composing the DP.

A standard NS is formed by a combination of VNFs/CNFs while a P4-based NS includes a P4NF, which describes the configuration of the custom DP dedicated to that NS. Therefore, a new workflow must be defined to accommodate the management of the lifecycle of these P4NFs. In this case, Juju is in charge of the lifecycle management of the P4NFs, being responsible for their instantiation and termination. Thus, a P4NF is a PNF that points to the physical P4 switch and implements the required Juju charms for its configuration.

To achieve these features, maintaining different types of state information is crucial. This leads to the following requirements:

- Minimal Layer 2 (L2) forwarding requires the P4 switch to know the MAC addresses connected to each physical port. Traffic from different VNFs/CNFs could come into the same switch port as they could be hosted by the same compute node. A static match-action table (MAT), *forward_l2*, is used for this purpose. Each compute node SR-IOV VF is preconfigured with a static MAC address, and the entries with each VNF/CNF MAC address to switch port mapping are preloaded into the aforementioned table.
- Multi-tenancy and DP slicing support require maintaining a record in an external database of which DP functions are in use by each NS. This information is stored in the *ns_functions* table, which is consulted whenever the DP needs to be recomposed.
- Upon the instantiation or termination of a P4NF that requires recompiling the DP, all CP information is lost as the DP tables are destroyed. For this reason, an image of the current CP state must be kept up to date in an external database. The *cp_rules* table is introduced.
- The P4 code used by any P4NF needs to be stored in a public repository. This includes the code that would be required by a new P4NF, in order to pull said code during the instantiation of the P4NF. This repository, named *P4 function repository*, allows any user to upload their custom DP functions so they are beforehand checked.

When an NS is instantiated, VNFs and CNFs are configured following the standard lifecycle operations defined in OSM. For the P4NF, the code defined in the Juju charms performs the following lifecycle operations to instantiate the P4NF:

1) **Check required P4 functions**: The P4NF metadata file, which is part of the Juju charms, includes a list of the P4 functions that the DP must implement to support the P4NF.
2) **Update *ns_functions* table**: As the P4 switch cannot run more than a single DP P4 program at once, this database table is required in order to obtain the complete list of the running P4 functions and recreate the program in the following step. Likewise, the table must be updated to reflect the DP pipeline state after the introduction of the new P4NF.
3) **Compose P4 program**: This step consists in combining different P4 functions to create a single P4 program that configures the DP pipeline. Even if the new NS only requires P4 functions that already exist in the DP, a new DP needs to be composed, as the tables must be unique to restrict access to each DP table from the CP. This process is challenging and requires specific P4 coding rules to achieve composable P4 functions. The literature provides different approaches to accomplish this task [5], [6]. To use custom P4 functions, they must be uploaded to the *P4 function repository*, which assures their ”composability”.
4) **Compile P4 program**: The Juju charms establish an active connection to the P4 switch to load, compile, and run the P4 program. During this step, the P4 switch is not operational until the new program is running.
5) **Configure switch ports**: in case the DP was recompiled, it is necessary to bring up all the ports that are in use by the switch. This is a fixed task, as all the ports with a physical connection to a compute node need to be enabled.
6) **Update CP**: Finally, the Juju charms perform a basic CP configuration to provide L2 connectivity between the VNFs/CNFs specified in the NS. This is achieved by having the VNFs/CNFs send their MAC addresses to the P4NF through Juju. These parameters are then used as input parameters for the CP to configure the *forward_l2* MAT. On the other hand, the *cp_rules* database table needs to be consulted to restore the most recent state of the CP.

Additionally, CP operations that involve filling owned DP tables are allowed through the NF day-2 lifecycle operations supported by OSM and Juju. This includes both setting up the initial CP rules after the P4NF instantiation, and inserting new rules on demand. The exception to this is the *forward_l2* MAT, which, despite containing entries for all users, is managed by the network administrator only. Each day-2 operation is registered in the *cp_rules* database to keep it up to date.

The termination of the NS must be considered as part of the lifecycle management. Therefore, the code defined in the charms performs the following operations to remove the P4NF:

1) **Update *ns_functions* table**: All active P4 functions are retrieved from the *ns_functions* table. Those functions that are no longer needed after the termination of the P4NF are removed from the table.
2) **Compose P4 program**: Similarly to the NS instantiation, the P4 program is reconfigured. The no longer needed P4 functions after the termination of the P4NF are excluded to build the new P4 program. This includes the removal of unused MATs. This process is done to achieve the simplest possible DP pipeline.
3) **Compile P4 program**: The same as in the NS instantiation workflow is applied.
4) **Configure switch ports**: The same as in the NS instantiation workflow is applied.
5) **Update CP**: The entries in the *forward_l2* MAT for the terminated P4NF are discarded. In addition, CP changes triggered through day-2 operations need to be reverted during the reconstruction of the CP. This is done by looking up the *cp_rules* table.

IV. CONCLUSIONS AND FUTURE WORK

DP acceleration in NFV is a promising idea, as the dynamic reconfiguration of the DP in an NFV-based environment grants network service providers with powerful possibilities. However, there are still some challenges that must be overcome in order to optimize the system for production environments. We hereby highlight and discuss the following three challenges.

The first challenge is sharing P4 MATs among tenants. This might seem a good option, as it results in better usage of resources and optimization of the P4 program. However, if a table is shared between two tenants, one could change the entries that correspond to the other tenant and illegitimately modify its behavior. In fact, unauthorized changes in the communication among VNFs could be a potential attack vector inside the system. The proposed solution has been to duplicate the tables for each tenant or NS, so that each of them only has access to its own tables and cannot modify other equivalent tables.

The second challenge is concerned with the duplication of parser code in P4 programs. The P4 program in the switch implements P4 functions in a modular way, only composing and compiling P4 functions required by instantiated NSs. However, multiple NSs may have partially concurrent needs, or, in other words, share one P4 function but require others that are different among them. This leads to different parsing requirements of those NSs. To put this in context, while two NSs might use the same static L2 forwarding table, one NS may just need to parse the Ethernet header, and the other one may need to parse even the application layer —in order to execute other actions in a later phase. Therefore, a solution to this could be the duplication of parsers —one for each NS—, or having a "super-parser" that is able to fulfill any parsing requirement of all instantiated NSs, using only the necessary parts for each case.

The third challenge is related to the service downtime that arises due to the recompilation of the P4 program during the instantiation and deletion periods of NSs. Each time one of these processes is executed, the DP must be recomposed and recompiled, interrupting the service in the compilation part. This means that all deployed NSs lose connectivity in that period, which entails an availability issue. Approaches similar to [5], [7] could be studied to address this issue, in which the switch enters in a "slow-mode" state while the new DP is being recompiled, but does not stop serving with the current DP. Even though experiencing service downtime might not be reasonable for critical services, it could be for experimental environments that are not significantly affected by it.

Therefore, our research shows that there are several methods under investigation that can provide solutions to the aforementioned challenges, and advance towards a fully operational system. There is still more work to do, mainly focusing on the challenges analyzed in this section, but we think that the overall approach is promising and may change the current service deployment system in NFV environments.

ACKNOWLEDGEMENTS

This work was supported in part by the Spanish Ministry of Science and Innovation through the national project (PID2019-108713RB-C54) titled "Towards zeRo toUch nEtwork and services for beyond 5G" (TRUE-5G), and in part by the "Smart Factories of the Future" (5G-Factories) (COLAB19/06) project.

REFERENCES

[1] Mafiolettik, Diego Rossi, et al. "PIaFFE: A Place-as-you-go In-network Framework for Flexible Embedding of VNFs." ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.
[2] Bonofiglio, Gaetano, et al. "Kathará: A container-based framework for implementing network function virtualization and software defined networks." NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2018.
[3] He, Mu, et al. "P4nfv: An nfv architecture with flexible data plane reconfiguration." 2018 14th International Conference on Network and Service Management (CNSM). IEEE, 2018.
[4] Stoyanov, Radostin, and Noa Zilberman. "MTPSA: Multi-Tenant Programmable Switches." Proceedings of the 3rd P4 Workshop in Europe. 2020.
[5] Soni, Hardik, et al. "Composing dataplane programs with μP4." Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication. 2020.
[6] Zaballa, Eder Ollora, et al. "A perspective on P4-based data and control plane modularity for network automation." Proceedings of the 3rd P4 Workshop in Europe. 2020.
[7] Osiński, Tomasz, et al. "DPPx: A P4-based Data Plane Programmability and Exposure framework to enhance NFV services." 2019 IEEE Conference on Network Softwarization (NetSoft). IEEE, 2019.
[8] Osiński, Tomasz, et al. "Offloading data plane functions to the multi-tenant Cloud Infrastructure using P4." 2019 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS). IEEE, 2019.
[9] ETSI ISG NFV, "ETSI GS NFV-IFA 001 V1.1.1: Network Functions Virtualisation (NFV); Acceleration Technologies; Report on Acceleration Technologies Use Cases." 2015.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

An NFV system to support service provisioning on UAV networks

Borja Nogales[1], Iván Vidal[1], Victor Sánchez-Aguero[1,2], Francisco Valera[1], Luis F. Gonzalez[1]

[1] Telematic Engineering Department, Universidad Carlos III de Madrid, Avda. Universidad, 30, 28911 Leganés (Madrid), Spain.

[2] IMDEA Networks Institute. Avda. del Mar Mediterráneo, 22, 28918 Madrid, Spain.

bdorado@pa.uc3m.es, ividal@it.uc3m.es, victor.sanchez@imdea.org, fvalera@it.uc3m.es, luisfgon@it.uc3m.es

I. Abstract

In this presentation, we will first describe the design and implementation of an NFV system capable of deploying moderately complex network services over a wireless ad-hoc network of resource-constrained compute nodes. The system design targets aerial networks built by Unmanned Aerial Vehicles (UAVs), and it relies on container virtualization to support the execution of network functions within constrained environments, as well as on mobile ad-hoc networking to support the underlying end-to-end network communications [1]. The presentation will also cover the implementation experience from developing this NFV system, which is based on relevant and widely-adopted open-source technologies in the NFV arena such as ETSI Open-Source MANO (OSM) and OpenStack.

In addition, we will present the details concerning the integration of this system into a distributed NFV testbed spanning three different remote sites in Spain, i.e., Universidad Carlos III de Madrid (UC3M), Universidad Politécnica de Cataluña (UPC), and Universidad del País Vasco (UPV-EHU). The goal of this testbed is to explore synergies among NFV, UAVs, and 5G vertical services, following a practical approach primarily governed by experimentation. To showcase the potential of this testbed to support vertical services, we will present three different use cases that have been realized as part of our prior research work: *i)* the automated deployment of an IP telephony service on a delimited geographic area, using a network of interconnected UAVs [2] (noteworthily, this work was awarded by ETSI as the best proof-of-concept demonstration with OSM during the OSM Release Eight cycle [3]); *ii)* the realization of a smart farming vertical service [4]; and *iii)* a public-safety vertical use case, which uses aerial and vehicular NFV infrastructures to monitor traffic conditions and handle emergency situations [5]. This latter involves an international collaboration with the Instituto de Telecomunicações of Aveiro, which operates a vehicular NFV infrastructure.

Finally, the presentation will tackle the standardization challenges related with the future view of a decentralized and flexible MANO framework, capable of supporting the operation of cost-effective, reliable services beyond the edge of the telecommunication operator infrastructures. In this view, multiple stakeholders would collaboratively provide a wide range of heterogeneous compute-connect devices (e.g., end-user terminals, CPEs, or UAV swarms). These devices might exist and be opportunistically used, or they could otherwise be deployed on-demand by those stakeholders, contributing to the availability of a potentially unlimited pool of network, computing, and storage resources beyond the network edge. This view introduces several standardization challenges to the NFV MANO framework in terms of interoperation, flexibility, robustness, and security. These challenges have been presented at the *NFV Evolution*[1] event organized by ETSI, and will build the basis of our future work in this research line.

II. Acknowledgment

This work has been partially supported by the European H2020 LABYRINTH project (grant agreement H2020-MG-2019-TwoStages-861696), and by the TRUE5G project (PID2019-108713RB-C52PID2019-108713RB-C52/AEI/10.13039/501100011033) funded by the Spanish National Research Agency.

III. References

[1] B. Nogales, V. Sanchez-Aguero, I. Vidal, and F. Valera, "Adaptable and Automated Small UAV Deployments via Virtualization.", *Sensors*, vol. 18, no. 12, p. 4116, Nov. 2018.

[2] B. Nogales, I. Vidal, V. Sanchez-Aguero, F. Valera, L. Gonzalez, and A. Azcorra, "Automated Deployment of an Internet Protocol Telephony Service on Unmanned Aerial Vehicles Using Network Functions Virtualization.", *JoVE* Vis. Exp. (153), e60425.

[3] OSM. "OSM PoC 10 Automated Deployment of an IP Telephony Service on UAVs using OSM.", [Online] Available: https://osm.etsi.org/wikipub/index.php/OSM_PoC_10_Automated_Deployment_of_an_IP_Telephony_Service_on_UAVs_using_OSM, Accessed on: Jun. 16, 2021.

[4] I. Vidal, B. Nogales, F. Valera, L. F. González, V. Sanchez-Aguero, E. Jacob, and C. Cervelló-Pastor, "A Multi-Site NFV Testbed for Experimentation With SUAV-Based 5G Vertical Services.", *IEEE Access*, 8, 111522-111535, 2020.

[5] B. Nogales, M. Silva, I. Vidal, M. Luís, F. Valera, S. Sargento, and A. Azcorra, "Using Aerial and Vehicular NFV Infrastructures to Agilely Create Vertical Services.", *Sensors*, 21(4), 1342, 2021.

[1] ETSI *NFV Evolution* event. All details included in: https://www.etsi.org/events/1892-nfvevolution


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Optimización Adaptativa basada en Colonias de Hormigas para la Composición de Cadenas de Funciones Virtuales en una Red 5G Dinámica

Segundo Moreno, Antonio M. Mora
Dto. Teoría de la Señal, Telemática y Comunicaciones
ETSIIT-CITIC, Universidad de Granada, España
segundomoto@correo.ugr.es, amorag@ugr.es

***Resumen*—Las redes 5G dependen en gran medida de la gestión y el procesamiento basados en software. Las redes definidas por software (SDN) y la virtualización de funciones de red (NFV) forman parte del núcleo de estas. Los servicios ofrecidos dentro de este entorno se componen de varias funciones de red virtuales (VNF) que deben ejecutarse en un orden (normalmente) estricto. Esto se conoce como *Service Function Chaining* (SFC) y, dado que esas VNFs podrían estar ubicadas en diferentes nodos a lo largo de la red, además de la baja latencia esperada en el procesamiento de los servicios 5G, hace que el SFC sea un problema de optimización difícil de resolver. En un trabajo anterior, los autores presentaron un algoritmo de Optimización de Colonias de Hormigas (ACO) para la minimización del *coste de enrutamiento* de la composición de la cadena de servicios, tratándose de una aproximación preliminar capaz de resolver instancias simples y 'estáticas'; es decir, aquellas en las que la topología de la red permanece invariable durante la resolución. Esto dista mucho de la situación real de las redes, en las que normalmente los nodos (y los enlaces) aparecen y desaparecen continuamente. Así, en este trabajo describimos una evolución de nuestra propuesta anterior, que considera un modelo dinámico del problema, más cercano al escenario real. De manera que, en las instancias, los nodos y enlaces pueden ser eliminados o activados repentinamente. El algoritmo ACO será capaz de adaptarse a estos cambios y seguir ofreciendo soluciones óptimas. Dicho método ha sido probado en tres instancias dinámicas de diferentes tamaños, obteniendo resultados muy prometedores.**

***Palabras Clave*—Virtualización de Funciones de Red, NFV, Cadena de Funciones Virtuales, Enrutamiento, Routing, Redes 5G, Metaheurísticas, Algoritmos de Optimización basada en Colonias de Hormigas, OCH, ACO**

I. INTRODUCCIÓN

Las tecnologías de red actuales se centran en la enorme demanda que tienen estas, tanto en lo que respecta al número de dispositivos conectados a ellas como a los exigentes requisitos de baja latencia y gran ancho de banda. Las redes 5G han sido diseñadas para hacer frente a estas nuevas características siendo capaces de servir con flexibilidad a las necesidades de los usuarios y servicios.

Así, las redes 5G se desplegarán sobre nuevas tecnologías facilitadoras que permitirán la realización de una red virtualizada, programable y flexible. Dos de estas tecnologías son las redes definidas por software (SDN) y la virtualización de las funciones de red (NFV), que aún se encuentran en fase de desarrollo y evolución [1]. Las SDN pretenden separar el reenvío y el procesamiento del tráfico, basándose en la automatización de algunas operaciones de gestión de la red. La NFV se basa en la tecnología de virtualización para ejecutar funciones de red implementadas por software. Estas dos tecnologías se combinan en el llamado composición de cadenas de funciones de red, o *Service Function Chaining* (SFC), cuyo objetivo es establecer dinámicamente nuevos servicios de red a través de un conjunto de funciones virtuales, que deben ejecutarse en un orden específico [2].

Este trabajo aborda la composición óptima de estas cadenas de servicios. Así, dado un grafo de red en el que cada nodo puede ejecutar diferentes funciones virtuales; y considerando una cantidad de recursos por nodo, los recursos que requiere una función, el ancho de banda en los enlaces, y el orden de composición deseado; el objetivo es construir el camino óptimo y válido correspondiente a un servicio de red, que minimice el coste de encaminamiento (es decir, el número de saltos entre nodos).

Se trata de un problema de dificultad NP-completo [3], que los autores resolvieron con una primera aproximación basada en un algoritmo de optimización basada en colonias de hormigas (ACO) [4]. Los algoritmos ACO [5] se inspiran en el comportamiento de las hormigas naturales cuando buscan comida y se aplican para resolver problemas de optimización combinatoria, por lo que utilizan una colonia de *hormigas artificiales*, que son agentes computacionales que se comunican entre sí mediante una *matriz de feromonas*. Estos agentes trabajan sobre problemas formulados en un grafo con pesos en sus arcos. En cada iteración, cada hormiga construirá un

camino completo (solución) moviéndose a través de él. Una vez construido el camino (o durante su construcción), la hormiga depositará un rastro de feromona que, generalmente, estará relacionado con la bondad de la solución. Por lo tanto, este rastro será una medida (informativa para otros) de lo deseable que es seguir la misma ruta que la citada hormiga.

Así, desarrollamos en [4] una variación de ACO bautizada como *Ant-SFC*, inspirada en el modelo de ACO más simple, el Sistema de Hormigas [6]. Sin embargo, los escenarios en los que se probó el algoritmo tenían una clara desventaja: la ausencia de dinamismo. Las redes son sistemas en continuo cambio, es decir, los nodos (y también los enlaces) surgen o desaparecen constantemente. Este comportamiento por ello se incorpora a las instancias para resolver el SFC de manera más fidedigna.

Para ello, el presente trabajo mejora tanto la definición y modelización del problema -haciéndolo más cercano a la realidad-, como el algoritmo para abordarlo. De modo que se han considerado instancias dinámicas, en las que pueden ocurrir algunos eventos, por lo que nodos o enlaces pueden activarse o desactivarse repentinamente. Por tanto, el algoritmo se ha transformado en lo que hemos denominado *Dynamic Ant-SFC (DAnt-SFC)*, capaz de adaptar su comportamiento para encontrar soluciones óptimas en estos escenarios cambiantes.

Se han definido nuevas instancias del problema para probar el algoritmo propuesto, incluyendo todas ellas los eventos mencionados; también se ha considerado un escenario más complejo, teniendo este 52 nodos. En los experimentos se ha analizado la capacidad de adaptación del algoritmo y su rendimiento en la reconstrucción de soluciones óptimas, así como la habilidad de recuperación que se ofrece a la red en esos eventos concretos.

II. PROBLEMA A RESOLVER: ENRUTAMIENTO DINÁMICO DE COSTE MÍNIMO PARA SFC

Este trabajo se centra en la composición de una cadena de funciones de servicio (SFC) en una red. El proceso de composición de la SFC es uno de los principales retos de NFV, ya que en esta tarea intervienen tanto el cálculo de la ruta como la dirección del tráfico. Además, debido a las propiedades de SFC, las rutas de flujo se definen como un conjunto ordenado en cadena de funciones de servicio o *Service Functions* (SF) que maneja el tráfico de la entrega, el control y la supervisión de un servicio/aplicación específico [7].

En este entorno y con el fin de mejorar el rendimiento y ahorrar recursos en la red, se requiere una estrategia óptima. Así, hemos bautizado este problema como *Optimización del Enrutamiento para SFC (OR-SFC)*. En este proceso, será necesario determinar el camino que deben seguir los datos entre las funciones de red virtuales adyacentes para cada uno de los servicios solicitados. La Figura 1 muestra un ejemplo de instancia de este problema.

En él, la petición del usuario se denomina *conexión*, y está definida por una tupla, *C=(origen, destino, valor de la demanda, [funciones a ejecutar])*, en la que la conexión tiene como origen el nodo '1' y como destino el nodo '6', con una demanda de tráfico de 2 Mbps y las funciones virtuales a ejecutar 3, 5 y 6, en ese preciso orden. Si nos fijamos en la figura, el valor cercano a los enlaces hace referencia al correspondiente ancho de banda disponible en cada uno. En cada nodo se han indicado las funciones de red que puede ejecutar (cubos). Además, cada uno de los nodos tendrá asociados unos recursos informáticos (CPU, Memoria, espacio en disco) agrupados en un único número por simplicidad. Del mismo modo, en la definición del problema habrá una lista que asocie un coste en recursos a cada función, como sus requisitos para ser ejecutada.

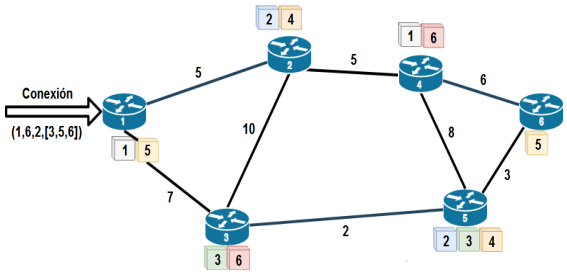


Fig. 1. Ejemplo de problema de Routing para SFC

En el ejemplo (Figura 1), una solución debe servir las funciones 3, 5 y 6 (en este orden estricto). Así, una posible solución podría ser el camino $[1 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 4 \rightarrow 6]$, pero también $[1 \rightarrow 3 \rightarrow 5 \rightarrow 6 \rightarrow 4 \rightarrow 6]$. La elección de uno u otro afectará al coste del encaminamiento en la red, así como a su rendimiento global. La ruta también debe cumplir algunas restricciones adicionales, como que los enlaces deben tener suficiente ancho de banda para cubrir la demanda, y los nodos deben tener suficientes recursos disponibles para ejecutar las funciones virtuales.

Además de lo descrito anteriormente, se añade al problema un *componente dinámico*. Tratará de simular un comportamiento más realista de la red, manejando posibles cambios o modificaciones durante la resolución del problema.

La forma en que se ha implementado esta parde dinámica añade al problema una componente de complejidad, ofreciendo resultados más realistas. Esta posibilidad ofrecerá potencialmente una gran flexibilidad a la hora de gestionar esta red dentro de un entorno virtualizado, ya que en la gran mayoría de los casos podrá hacer frente a situaciones imprevistas al no requerir que la red permanezca estática en cuanto a su estructura, llevando el modelo del problema a un enfoque más realista, considerando que las redes reales son completamente dinámicas.

III. ESTADO DEL ARTE

SFC es uno de los principales retos en NFV. Este debe ser tratado como un problema de optimización NP-completo. Es por ello que ha atraído la atención de la academia, proponiendo diferentes soluciones para resolverlo, que principalmente se centran en soluciones exactas o heurísticas, siendo sólo algunas de las propuestas

las que utilizan métodos de inteligencia computacional avanzada, como las metaheurística.

En cuanto a las aproximaciones exactas para resolver el problema OR-SFC, la mayoría de ellas se centran en modelos de optimización basados en técnicas de Programación Lineal. Por ejemplo, los autores en [2] presentaron un modelo que resuelve el enrutamiento SFC y la asignación de funciones virtuales para los intervalos de horas pico. Cabe destacar también el trabajo desarrollado en [8], en el que los autores formularon un modelo matemático que resuelve el problema utilizando métodos de descomposición.

También existen algoritmos heurísticos para resolver este tipo de problemas. Estas heurísticas son útiles para encontrar soluciones factibles y precisas para instancias mayores del problema en un tiempo de cálculo reducido. Los algoritmos *Greedy* son muy utilizados para este fin, como se hace en [9] y [10]. En general, se ha demostrado que las heurísticas producen soluciones aproximadas cercanas al óptimo y son apropiadas cuando hay que resolver instancias grandes. En estos casos, la heurística proporciona un equilibrio óptimo entre la validez de la solución y el coste de cálculo [11].

Por el contrario, las metaheurísticas no se han aplicado demasiado en este ámbito. Aunque el problema OR-SFC es muy adecuado para la técnica de Optimización basada en Colonias de Hormigas si bien, hasta donde sabemos, sólo nuestro trabajo anterior [4] se ha centrado en la resolución de este problema aplicando ACO. Otros enfoques en la literatura se enfrentan en cambio a la llamada Asignación de Recursos en NFV, como [12] donde los autores aplican Tabu Search; o [13] en el que los autores abordan la optimización de la ubicación de Funciones Virtuales de Red (VNFs) en los nodos de la red, considerando el consumo de energía en los servidores requeridos, aplicando una variación del Algoritmo Genético.

Dentro de este ámbito, algunos trabajos tratan de resolver problemas análogos aplicando un enfoque desde un punto de vista dinámico, donde la topología de la red puede cambiar. Este es el caso de [14] donde se adoptan diferentes colonias de hormigas simultáneamente para favorecer la exploración en redes dinámicas, evitando en gran medida el estancamiento. Otro caso se da en [15], donde se utiliza un algoritmo basado en ACO para resolver el enrutamiento dinámico anycast y la asignación de longitudes de onda en redes ópticas, ofreciendo reducciones de la probabilidad de bloqueo y mejoras sobre otros métodos. Sin embargo, de nuevo, ninguna de las propuestas se centra en la resolución del problema SFC.

IV. ALGORITMO IMPLEMENTADO: ANT-SFC DINAMICO

El algoritmo implementado para abordar la versión dinámica del problema es una 'evolución' de nuestro anterior Ant-SFC [4], es decir, una adaptación del Sistema de Hormigas clásico [6] para resolver el problema SFC estándar. Así, cubrirá la resolución de caminos óptimos en un modelo de red de telecomunicaciones, cumpliendo con las siguientes restricciones (ver Sección II):

- Un camino debe ser definido en el grafo que modela la red para cada resolución de conexión (solicitud de servicio). Debe pasar por nodos disponibles que puedan servir cada una de las funciones de red requeridas, en el orden dado (indicado por cada servicio).
- Cada enlace debe tener suficiente capacidad (ancho de banda disponible) para poder satisfacer la demanda de tráfico de cada conexión. Ésta disminuirá cada vez que una ruta pase por un enlace.
- Los nodos, al igual que los enlaces, deben tener suficientes recursos disponibles para la ejecución de las funciones requeridas. Los recursos de los nodos disminuirán según la demanda de cada función ejecutada.

Estas restricciones deben cumplirse incluso teniendo en cuenta que los nodos o enlaces podrían activarse o desactivarse repentinamente, ya que las instancias del problema son dinámicas. Por lo tanto, el algoritmo adaptado se ha denominado *Dynamic Ant-SFC* o simplemente DAnt-SFC.

El *dinamismo* implementado pretende añadir un componente aún más realista a las redes simuladas. Se conseguirá mediante la eliminación o introducción de nodos y/o enlaces en determinados momentos de la ejecución, haciendo que el algoritmo sea capaz de recuperarse de eventos de potencial colapso de nodos principales. De este modo, el algoritmo recibirá algunas tuplas dentro de un archivo de eventos, como una modelización "controlada" y simplificada de ese dinamismo.

El formato de cada tupla es *(número de iteración, activación-desactivación del nodo, nodo, activación-desactivación del enlace, nodo origen del enlace, nodo destino del enlace)*. Por ejemplo, la tupla "(3, 0, 4, 0, 4, 8)" equivaldría a decirle al algoritmo que a partir de la iteración 3, el nodo 4 se desactivará junto con el enlace de 4 a 8, también desactivado e inutilizable, a menos que se reactive explícitamente de nuevo en un evento posterior.

Cada vez que una de estas tuplas sea procesada por la parte de dinamismo, se activará automáticamente una parte del algoritmo para eliminar o añadir el nodo correspondiente, asegurando así que las sucesivas soluciones que se calculen considerarán una versión actualizada de la red. Además, al final de cada proceso, se comprueba si existe una solución potencialmente mejor, teniendo en cuenta el tiempo y la capacidad de los nodos probados.

La resolución de cada *conexión* (solicitud de servicio) se ha planteado como una búsqueda de camino óptimo individual, aunque sean dependientes entre sí por el consumo de ancho de banda de los enlaces y de recursos de los nodos. El cuerpo principal de DAnt-SFC se presenta en el Algoritmo 1.

La adaptación del algoritmo se ha centrado en los siguientes aspectos:

- *Inicialización del grafo*: el grafo tiene que estar inicializado tal y como estaba antes de que la hormiga anterior lo modificara construyendo su solución antes de que cualquier hormiga comience a construir una solución.

Algorithm 1 DAnt-SFC ( )

Algoritmo principal DAnt-SFC

```
Inicializacion_parametros()
Leer_configuracion_red()
Leer_conexiones()
Leer_configuracion_dinamismo()
/* Se buscará una solución por cada conexión */
for cada conexión c do
    while criterio_finalizacion_no_terminado do
        for cada hormiga h do
            s[h]=Construir_Solucion(c,h)
        end for
        /* En todos los enlaces del grafo */
        Evaporacion_Feromona()
        /* Links usados por la mejor hormiga */
        s*=Elegir_Mejor_Solucion(s[h])
        Actualizacion_Feromona_Global(s*)
        Actualizar_estado_dinamismo() /* Comprueba si la mejor solución es
        todavía válida*/
    end while
    /* Ancho de banda de los enlaces y los nodos son actualizados */
    Actualizacion_Red(c,s*)
end for
```

- *Heurística*: se ha considerado asignar una mayor probabilidad de ser elegido a los enlaces con mayor ancho de banda disponible, intentando minimizar el riesgo de agotamiento de los enlaces. Además, se ha incluido una condición para la selección del siguiente nodo en la construcción de la solución: si alguno de los nodos es capaz de servir a la siguiente función de red que espera ser servida en la cadena, se duplicará la probabilidad de elección del nodo para guiar a la hormiga hacia él.
  Obviamente, el enlace debe tener suficiente ancho de banda disponible para cada caso y el nodo seleccionado tiene que tener suficientes recursos para ejecutar la siguiente función de red en la cadena.
- *Ruleta de probabilidades*: se utilizará una ruleta de probabilidades como política de decisión para el siguiente estado una vez asignada la probabilidad de pasar a cada nodo desde el real en la construcción de una solución. Esta ruleta consiste en asignar un espacio proporcional a la probabilidad de cada nodo en una "ruleta virtual" y su giro aleatorio para obtener el siguiente nodo elegido.
- *Restricción del ancho de banda de los enlaces*: se basa en la construcción de la lista de nodos factibles.
- *Restricción de recursos de nodos*: se basa en la construcción de la lista de nodos factibles.
- *Actualización de enlaces y nodos (construcción de una solución)*: el ancho de banda de los enlaces y los recursos de los nodos (en caso de que el nodo cumpla una función de red) se actualizan cada vez que una hormiga se desplaza hacia un nodo de la red mientras se construye una solución para resolver una conexión. Ambos valores se actualizan con la demanda de tráfico de la conexión y el coste en recursos de la función de red, respectivamente, evitando la generación de bucles infinitos.
- *Actualizaciones de la red (conexiones)*: la red se actualiza teniendo en cuenta la trayectoria definida por la solución cada vez que se encuentra una solución para una determinada cadena. Por lo tanto, los enlaces y nodos utilizados para este camino se actualizan siguiendo el método utilizado en el caso anterior.
- *restricción de trayectoria completa*: una solución sólo se considerará válida si comienza y termina en los nodos exactamente dados por la conexión. También tiene que pasar por los nodos que cumplen funciones de red en el orden dado por la propia conexión. Si no es así, la solución no se utilizará.
- *Coste de la ruta (conexión)*: será el *número de saltos* requerido en el grafo para la composición de la cadena de funciones necesaria para completar una conexión.
- *Coste de la solución global*: una solución completa estará formada por unos caminos de coste mínimo para resolver cada una de las conexiones solicitadas para la instancia seleccionada en un determinado intervalo de tiempo. Por lo tanto, el coste de la solución global será el número total de saltos de todas las conexiones solicitadas.
- *Actualización de la feromona*: al igual que el correspondiente *evaporación de la feromona* realizado en todos los enlaces tras la construcción de todas las soluciones, sólo se realizará una *actualización de la feromona* en los enlaces que formen parte de la mejor solución, que será directamente proporcional al número de saltos en esta misma. Cuanto menor sea el número de saltos, mayor será la contribución en los enlaces.

V. Experimentos y resultados

Se han considerado tres instancias diferentes para probar el algoritmo propuesto:

Instancia de 6 nodos: se utiliza como un enfoque conceptual, donde el algoritmo ha sido evaluado y validado de una manera más intuitiva. Las características del gráfico se pueden ver en: https://doi.org/10.6084/m9.figshare.14572200.v1 incluyendo las funciones de red disponibles en cada nodo así como el ancho de banda asociado a cada enlace. En esta instancia se consideran 3 conexiones diferentes a resolver, concretamente (ver formato en la sección II): Conexión 1: (A, F, 2, [3,5,6]), Conexión 2: (A, E, 8, [1,2,4]), y Conexión 3: (A, D, 5, [2,4,5]).

Instancia de 19 nodos: es un gráfico de 19 nodos que modela un caso más realista, más cercano a los que se resolverán en la realidad. La topología de esta instancia se puede consultar desde https://doi.org/10.6084/m9.figshare.14572224.v1. Las propiedades de los enlaces en esta instancia se pueden ver en https://doi.org/10.6084/m9.figshare.14572080.v1. Esta instancia está considerando estas 5 conexiones siguientes para resolver: Conexión 1: (H, J, 8, [5,1,2]), Conexión 2: (B, D, 8, [4,3,1]), Conexión 3: (Q, B, 1, [2,3,1]), Conexión 4: (R, J, 3, [5,2,3]), Conexión 5: (J, S, 8, [4,1,3]).

Instancia de 52 nodos: es un grafo de 52 nodos que modela un caso próximo a la realidad. Es la instancia más grande utilizada en este trabajo. Su topología se muestra en https://doi.org/10.6084/m9.figshare.14572230.v1, mientras

Algorithm 2 Construccion_Solucion (conexion, ant_id)

Algoritmo de construccion de una solucion DAnt-SFC

```
Inicializacion_hormiga(ant_id)
Inicializacion_red() /* Establece los valores de la red */
Aplicacion_dinamismo() /* Comprueba los nodos y enlaces activados/desactivados */
nodo_actual = conex.nodo_inicial
funcion_actual = conex.funciones[inicio]
L = guardar(nodo_actual) /* Lista de estados visitados */
F = guardar(funcion_actual) /* Lista de funciones servidas */
while (nodo_actual ≠ conex(nodo_final)) AND (funcion_actual ≠ conex.funcion[fin]) do
    /* A: lista nodos alcanzables, P: probabilidad de moverse a cada nodo alcanzable, Ω: restricciones del problema */
    P = calcular_probabilidades_transicion(nodo_actual, A, F, L, Ω)
    siguiente_nodo = ruleta_probabilidad(P, Ω)
    /* Actualización del ancho de banda de los enlaces */
    Actualizacion_Enlace(siguiente_nodo)
    L = guardar(siguiente_node)
    nodo_actual = siguiente_nodo
    /* Si la función está disponible será servida, y los recursos del nodo, actualizados */
    if funcion_actual in funciones.actuales_nodo[] then
        Actualizar_Nodo(funcion_actual)
        F = guardar(funcion_actual)
        funcion_actual = conex.siguientes(funciones[])
    end if
end while
```

que las tablas correspondientes a los enlaces de esta instancia pueden consultarse en https://doi.org/10.6084/m9.figshare.14483592.v2. Hay que resolver 10 conexiones: Conexión 1: (AP, K, 16, [3,1,2]), Conexión 2: (P, O, 6, [2,3,1]), Conexión 3: (R, AU, 11, [4,1,2]), Conexión 4: (AT, E, 5, [3,2,1]), Conexión 5: (AF, AE, 5, [1,3,2]), Conexión 6: (AA, AD, 14, [4,3,1]), Conexión 7: (S, I, 11, [4,2,1]), Conexión 8: (X, AD, 10, [3,2,1]), Conexión 9: (K, R, 7, [1,3,2]), Conexión 10: (AG, AX, 18, [3,1,2]).

A. *Resultados obtenidos*

Para la ejecución del algoritmo se ha utilizado un ordenador personal con un procesador Intel Core i5-1135G7 de 4 núcleos y 8 hilos a 2,40GHz, con 8GB de RAM DDR-4 y Windows 10 O.S. de 64 bits.

Las configuraciones del algoritmo consideradas en cada instancia se muestran en la Tabla I.

Tabla I
PARAMETROS CONSIDERADOS EN LOS EXPERIMENTOS

Parametro	Instancia 6N	Instancia 19N	Instancia 52N
$Iteraciones$	6	19	52
$Hormigas$	12	38	104
α (peso feromona)	1.2	1.2	1.2
β (peso heuristica)	2.0	2.0	2.0
ρ (factor evaporación)	0.3	0.3	0.3

Los valores dados se han fijado en base a las recomendaciones leídas en artículos sobre aplicación de ACO, como feromona y pesos heurísticos para el cálculo de la probabilidad de elección del siguiente nodo, aunque posteriormente se han ajustado y modificado siguiendo un proceso de experimentación sistemática. Los números de iteraciones y conexiones se han fijado con el fin de obtener buenas soluciones en un tiempo aceptable, aunque se podría llevar a cabo una investigación posterior sólo para determinar de forma óptima estos valores iniciales, pero no es el objetivo de este trabajo.

Dado que se trata de un algoritmo no determinista, para obtener resultados fiables, se han realizado 10 ejecuciones independientes resolviendo la misma instancia del problema (con el mismo número de conexiones) para cada escenario posible y para cada una de las tres instancias a probar.

Como se describe en la sección IV, se ha desarrollado una función de dinamismo para este algoritmo. Ésta permite obtener las mejores soluciones posibles mientras ciertos nodos de la red pueden caer o estar en línea durante la ejecución (simulando un escenario real de red de telecomunicaciones) sin modificar el comportamiento básico del código ejecutado. Se utiliza para reforzar el ya de por sí buen rendimiento del algoritmo en un mayor número de situaciones y escenarios diferentes.

- Versión elimina_nodo: *un nodo crítico, que forma parte de la mejor solución, es eliminado de la mejor solución hasta el momento*: con este tipo de variante, el algoritmo se ve obligado a recalcular la mejor ruta obtenida hasta el momento, dado que un nodo crítico será eliminado de la misma, no siendo posible por tanto su utilización y teniendo que aplicar las utilidades de la función de dinamismo para seleccionar la mejor opción posible.
- Versión elimina_dos_nodos: *Se eliminan dos nodos críticos de la mejor solución hasta el momento*: será similar al caso anterior, pero con una dificultad añadida (habrá una mayor parte de la matriz de nodos no disponible para ser seleccionada), haciendo que la nueva selección de rutas sea más desafiante y "realista".
- Versión elimina_restaura_nodo: *Se elimina un nodo crítico de la mejor solución hasta el momento y luego se restaura*: para esta situación, se eliminará un nodo crítico seleccionado y, después de un cierto número de iteraciones, se restaurará, observando si el algoritmo vuelve a seleccionar la mejor solución previamente guardada o no (si se ha visto obligado a mejorarla).

En la Tabla II, se pueden ver los resultados obtenidos en las simulaciones, tanto en número de saltos de la mejor

solución y tiempo de ejecución, como en tiempo, valor medio y desviación estándar.

Tabla II
RESULTADOS DEL ALGORITMO DANT-SFC PARA LAS INSTANCIAS DE 6, 19 Y 52 NODOS. SE ESPECIFICA: MEJOR EJECUCIÓN, COSTE, TIEMPO, VALOR MEDIO Y DESVIACIÓN ESTÁNDAR O STANDARD DEVIATION (SD), OBTENIDOS PARA 10 EJECUCIONES EN CADA VERSIÓN DINÁMICA.

Instancia 6N					
Versión	**Ejec.**	**Coste**	**t(s)**	**Media**	**SD**
elimina_nodo	7	13	0.073	13.8	0.707
elimina_dos_nodos	-	-	-	-	-
elimina_restaura_nodo	2	12	0.070	12.2	1.41
Instancia 19N					
Versión	**Ejec.**	**Coste**	**t(s)**	**Media**	**SD**
elimina_nodo	2	17	0.283	17.4	0.707
elimina_dos_nodos	8	16	0.258	16.2	0.707
elimina_restaura_nodo	3	17	0.303	17.8	1.41
Instancia 52N					
Versión	**Ejec.**	**Coste**	**t(s)**	**Media**	**SD**
elimina_nodo	5	35	4.906	35.5	1.414
elimina_dos_nodos	7	35	4.796	35.4	1.414
elimina_restaura_nodo	2	37	4.953	38.6	2.121

Cabe destacar que, para la simulación de la instancia más pequeña (6 nodos), al ser utilizada como referencia para replicar el funcionamiento de las demás y debido a su reducido tamaño, la versión elimina_dos_nodos de la prueba (en la que se eliminan dos nodos fundamentales) no puede realizarse como tal, ya que el cálculo de ciertas rutas sería imposible. Sin embargo, su correcto funcionamiento para este caso puede verificarse en las instancias más grandes. Por este motivo, se han utilizado la versión elimina_nodo y la versión elimina_restaura_nodo, siendo esta última igualmente válida como ejemplo de funcionamiento, ya que se elimina un nodo y se reactiva, obligando al algoritmo a recalcular las rutas (como efectivamente hace).

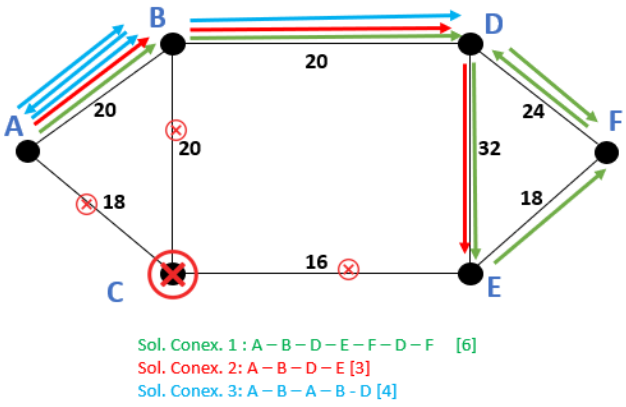


Fig. 2. Mejor solución encontrada para la Versión elimina_nodo (un nodo eliminado (C) en la iteración número 2) de la instancia de 6 nodos con 3 conexiones. Coste expresado en número de saltos junto a cada conexión. Figura con mayor resolución disponible en https://doi.org/10.6084/m9.figshare.16587497

Los resultados numéricos presentados en la Tabla II Se puede observar claramente la diferencia entre las distintas versiones ejecutadas. Para el caso de 6 nodos, en la primera versión, como se ve en la Figura 2, un nodo fundamental (C) bajará en la iteración número 2. Cuando

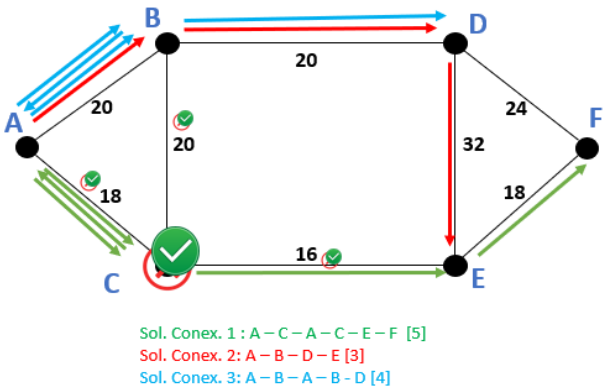


Fig. 3. Mejor solución encontrada para la Versión elimina_restaura_nodo (un nodo eliminado (C) en la iteración número 2, luego reactivado (C) en la iteración número 6) de la instancia de 6 nodos con 3 conexiones. Coste expresado en número de saltos junto a cada conexión. Figura con mayor resolución disponible en https://doi.org/10.6084/m9.figshare.16587512

este nodo cae, junto con sus enlaces, no puede ser utilizado en la red para futuros cálculos. En esta versión, esto ocurre casi desde el principio (iteración número 2), por lo que el algoritmo podría utilizar primero este nodo. Sin embargo, tras la correspondiente aplicación de la función de dinamismo, este nodo se desactivará y automáticamente, la mejor solución hasta ahora procesada, en caso de que lo contuviera, será eliminada y se obtendrá otra factible.

Se puede observar que para la Versión elimina_restaura_nodo, que se puede observar en la Figura 3, en la que este mismo nodo se cae pero en sucesivas iteraciones se reactiva de nuevo en la iteración número 6, el algoritmo detecta que es la ruta más eficiente y que puede volver a utilizar este nodo, ofreciendo además un menor coste para servir, por ejemplo, en la primera conexión requerida. De este modo, siempre se obtiene la mejor solución posible, teniendo en cuenta las capacidades que ofrece la red.

Como se ha mencionado anteriormente, con la instancia de 19 nodos se pueden observar gráficamente los cambios realizados en la red. En la primera versión, al poco de comenzar la ejecución, un nodo se cae de la red (nodo N) en la iteración número 4, lo que hace que el algoritmo recalcule las rutas en función de cómo ha quedado la topología. Después de todas las ejecuciones, el resultado es el que se muestra en la Figura 4. Esto podría ser un problema a priori, ya que uno de estos nodos fue utilizado en las rutas óptimas de la versión elimina_nodo. Sin embargo, la heurística acabará encontrando otra ruta óptima (de hecho, en este caso incluso mejor que la anterior) para encaminar la conexión según lo requerido y que se cumplan todos los requisitos.

Con respecto a la versión elimina_dos_nodos, ambos nodos N y M ya no están activos en la red (desconectados en la iteración 4 y 10 respectivamente), por lo que aparece una marca roja en ellos mismos y en sus enlaces. De nuevo, el algoritmo es capaz de recalcular nuevas rutas que no incluyan esa parte de la red sin mayor dificultad,

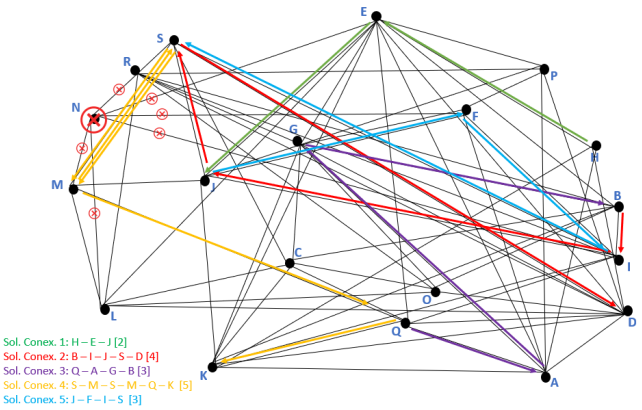


Fig. 4. Mejor solución encontrada para la Versión elimina_nodo (un nodo eliminado (N) en la iteración número 4) de la instancia de 19 nodos con 5 conexiones. Coste expresado en número de saltos junto a cada conexión. Figura con mayor resolución disponible en: https://doi.org/10.6084/m9.figshare.16587428

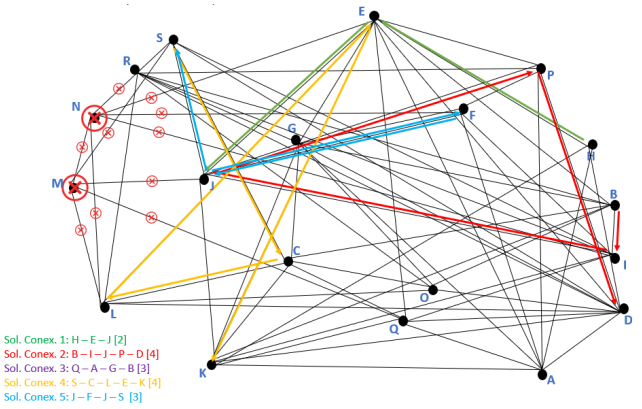


Fig. 5. Mejor solución encontrada para la Versión elimina_dos_nodos (dos nodos eliminados (M y N) en las iteraciones número 4 y 10 respectivamente) de la instancia de 19 nodos con 5 conexiones. Coste expresado en número de saltos junto a cada conexión. Figura con mayor resolución disponible en: https://doi.org/10.6084/m9.figshare.16587443

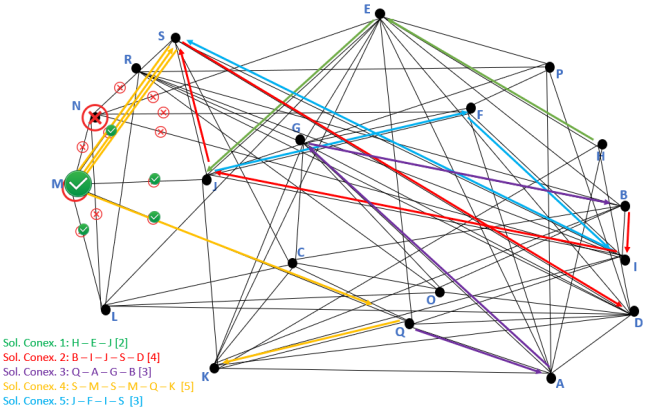


Fig. 6. Mejor solución encontrada para la Versión elimina_restaura_nodo (dos nodos eliminados (M y N) en las iteraciones 4 y 10 respectivamente, y luego uno reactivado (M) en la iteración número 16) de la instancia de 19 nodos con 5 conexiones. Coste expresado en número de saltos junto a cada conexión. Figura con mayor resolución disponible en: https://doi.org/10.6084/m9.figshare.16587446

como puede verse en la Figura 5.

Por último, los caminos formados en la Versión elimina_restaura_nodo para esta instancia de 19 nodos se pueden observar en Figura 6. Tras eliminar los nodos N y M en sucesivas iteraciones (iteración 3 y 9 respectivamente), las soluciones calculadas ahora no pueden incluirlos en las rutas obtenidas. Sin embargo, en la iteración número 16 (del total de 19 realizadas), el nodo M vuelve a estar activo, junto con sus correspondientes enlaces. Así, el algoritmo vuelve a incluirlo en sus tablas de nodos activos y lo tiene en cuenta para las nuevas soluciones; tanto es así que la solución final para la conexión 4 lo utiliza en sus resultados.

En la instancia de 52 nodos, se ha decidido mostrar gráficamente la Versión elimina_restaura_nodo de los experimentos realizados, mientras que de las versiones elimina_nodo y elimina_dos_nodos se mostrarán únicamente los resultados en la Figura 7. En este caso, se muestra cómo los nodos V y X se desactivan (en las instancias 4 y 8, respectivamente), y el sistema debe dejar de utilizarlos para el cálculo de las soluciones. Sin embargo, el nodo V vuelve a activarse a partir de la instancia 13, por lo que puede volver a utilizarse. De hecho, en la conexión 9, se utiliza para la mejor opción de la misma. Se puede observar la gran complejidad de la red y cómo el comportamiento del algoritmo es flexible y robusto ante los cambios, adaptándose a ellos en cada situación (Figura 8).

VERSION 1

Conex. 1: AQ – AP – AQ – L **[3]**
Conex. 2: Q – AZ – B – AZ – P **[4]**
Conex. 3: S – D – AM – AV **[3]**
Conex. 4: AU – AJ – F **[2]**
Conex. 5: AG – Q – AG – AL – AP – AF **[5]**
Conex. 6: AB – R – J – AE **[3]**
Conex. 7: T – O – AS – A – AQ – J **[5]**
Conex. 8: Y – X – T – AS – AE **[4]**
Conex. 9: L – S **[1]**
Conex. 10: AH – F – N – AI – AR – AY **[5]**

VERSION 2

Conex. 1: AQ – AP – AQ – L **[3]**
Conex. 2: Q – AZ – B – AZ – P **[4]**
Conex. 3: S – D – J – AV **[3]**
Conex. 4: AU – AX – F **[2]**
Conex. 5: AG – Q – AG – AL – AU – AF **[5]**
Conex. 6: AB – R – J – AE **[3]**
Conex. 7: T – O – AS – A – AQ – J **[5]**
Conex. 8: Y – Z – L – M – AE **[4]**
Conex. 9: L – S **[1]**
Conex. 10: AH – F – N – AI – AQ – AY **[5]**

Fig. 7. Resultados de la instancia de 52 nodos con 10 conexiones, Versión elimina_nodo y 2. Coste expresado en número de saltos junto a cada conexión.

VI. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo presenta una adaptación de un algoritmo de Optimización basada en Colonias de Hormigas o *Ant Colony Optimization*(ACO) para resolver el problema de enrutamiento en *Service Function Chaining* (SFC) dentro de una red definida por software (SDN). El problema se ha definido considerando también el dinamismo en la topología de la red, ya que los nodos y enlaces pueden ser activados o desactivados en cualquier momento.

El algoritmo propuesto se ha aplicado en tres instancias diferentes con distintos tamaños, y varios eventos de activación/desactivación predefinidos en algunos de los nodos y enlaces existentes.

A la vista de los resultados obtenidos, podemos concluir que *DAnt-SFC* es capaz de construir soluciones óptimas, incluso haciendo frente a cambios drásticos en la topología

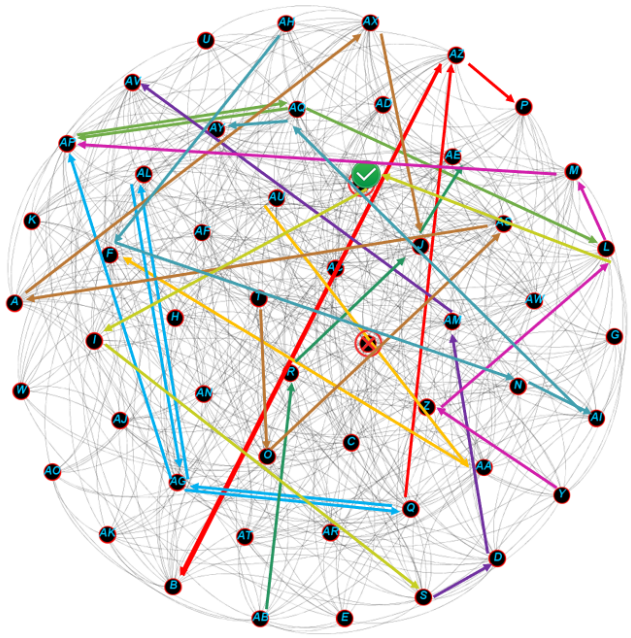


Fig. 8. Mejor solución encontrada para la Versión elimina_restaura_nodo (dos nodos eliminados (V y X) en las iteraciones 4 y 8 respectivamente), y luego uno reactivado (V) en la iteración 13) de la instancia de 52 nodos con 10 conexiones. Resultados de la Figura 9. Figura con mayor resolución disponible en https://figshare.com/s/b48a4c3c9f85ebca721d

Sol. Conex. 1: AQ – AP – AQ – L [3]
Sol. Conex. 2: Q – AZ – B – AZ – P [4]
Sol. Conex. 3: S – D – AM – AV [3]
Sol. Conex. 4: AU – AA – F[2]
Sol. Conex. 5: AG – Q – AG – AL – AG – AF[5]
Sol. Conex. 6: AB – R – J – AE [3]
Sol. Conex. 7: T – O – AS – A – AX – J [5]
Sol. Conex. 8: Y – Z – L – M – AF [4]
Sol. Conex. 9: L – V – I – S [3]
Sol. Conex. 10: AH – F – N – AI – AQ – AY [5]

Fig. 9. Resultados de la instancia de 52 nodos con 10 conexiones, Versión elimina_restaura_nodo, referida a la Figura 8. Coste expresado en número de saltos junto a cada conexión.

de la red como puede ser la caída de nodos críticos en las rutas previamente formadas, o la reincorporación de los mismos, debiendo recalcular las rutas formadas hasta dicho momento. Además, los tiempos de cálculo, siempre muy inferiores a 1 segundo para instancias pequeñas y medianas, aunque algo mayores (en torno a 5 segundos para las de mayor tamaño) son aceptables para la resolución de instancias consideradas ejecutadas en tiempo real.

En cualquier caso, una de las ventajas de los algoritmos ACO es su capacidad de obtener soluciones válidas desde la primera iteración, es decir, desde el mismo primer cálculo realizado durante una ejecución, así como su capacidad de autoadaptar su comportamiento a los cambios en la definición del problema, como sonlos cambios de topología en estas instancias.

Como trabajo futuro, probaremos mejores funciones heurísticas para guiar la construcción de soluciones en el algoritmo ACO. Además, se podrían implementar algunos enfoques híbridos, como los métodos de búsqueda local. También se probarán modelos ACO más sofisticados.

Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos RTI2018-102002-A-I00 (Ministerio de Ciencia, Innovación y Universidades), PID2020-113462RB-I00 (Ministerio de Ciencia e Innovación), TIN2017-85727-C4-2-P (Ministerio de Economía y Competitividad), B-TIC-402-UGR18 (FEDER y Junta de Andalucía), y el proyecto P18-RT-4830 (Junta de Andalucía).

Referencias

[1] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.

[2] V. Eramo, E. Miucci, M. Ammar, and F. G. Lavacca, "An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures," *IEEE/ACM Trans. Networking*, vol. 25, no. 4, pp. 2008–2025, 2017.

[3] T. Lukovszki, M. Rost, and S. Schmid, "It's a match!: Near-optimal and incremental middlebox deployment," *SIGCOMM Comput. Commun. Rev.*, vol. 46, pp. 30–36, Jan. 2016.

[4] S. Moreno, A. M. Mora, P. Padilla, J. Carmona-Murillo, and P. A. Castillo, "Applying ant colony optimization for service function chaining in a 5g network," in *Sixth International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2019, Granada, Spain, October 22-25, 2019* (M. A. Alsmirat and Y. Jararweh, eds.), pp. 567–574, IEEE, 2019.

[5] M. Dorigo and T. Stützle, "The ant colony optimization metaheuristic: Algorithms, applications, and advances," in *Handbook of Metaheuristics* (G. K. F. Glover, ed.), pp. 251–285, Kluwer, 2002.

[6] M. Dorigo, V. Maniezzo, and A. Colorni, "The ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 26, no. 1, pp. 29–41, 1996.

[7] A. M. Medhat, T. Taleb, A. Elmangoush, G. A. Carella, S. Covaci, and T. Magedanz, "Service function chaining in next generation networks: State of the art and research challenges," *IEEE Communications Magazine*, vol. 55, pp. 216–223, February 2017.

[8] N. Huin, B. Jaumard, and F. Giroire, "Optimal Network Service Chain Provisioning," *IEEE/ACM Transactions on Networking*, vol. 26, pp. 1320–1333, jun 2018.

[9] Z. Allybokus, N. Perrot, J. Leguay, L. Maggi, and E. Gourdin, "Virtual function placement for service chaining with partial orders and anti-affinity rules," *Networks*, vol. 71, no. 2, pp. 97–106, 2018.

[10] L. Qu, M. Khabbaz, and C. Assi, "Reliability-Aware Service Chaining in Carrier-Grade Softwarized Networks," *IEEE Journal Selec. Areas in Communications*, vol. 36, no. 3, pp. 558–573, 2018.

[11] T.-M. Nguyen, M. Minoux, and S. Fdida, "Optimizing resource utilization in NFV dynamic systems: New exact and heuristic approaches," *Computer Networks*, vol. 148, pp. 129–141, jan 2019.

[12] J. Gil-Herrera and J. F. Botero, "A scalable metaheuristic for service function chain composition," in *2017 IEEE 9th Latin-American Conference on Communications, LATINCOM 2017*, vol. 2017-Janua, pp. 1–6, Institute of Electrical and Electronics Engineers Inc., dec 2017.

[13] L. Laaziz, N. Kara, R. Rabipour, C. Edstrom, and Y. Lemieux, "FASTSCALE: A fast and scalable evolutionary algorithm for the joint placement and chaining of virtualized services," *Journal of Network and Computer Applications*, vol. 148, p. 102429, dec 2019.

[14] K. M. Sim and W. H. Sun, "Multiple ant-colony optimization for network routing," in *First International Symposium on Cyber Worlds, 2002. Proceedings.*, pp. 277–281, 2002.

[15] K. Bhaskaran, J. Triay, and V. M. Vokkarane, "Dynamic anycast routing and wavelength assignment in wdm networks using ant colony optimization (aco)," in *2011 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2011.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Auditoría Wi-Fi basada en placas de bajo coste

Anxo Otero, Carlos Dafonte, Diego Fernandez, Fidel Cacheda, Manuel López-Vizcaíno, Francisco J. Novoa
Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC)
Facultad de Informática
Departamento de Ciencias de la Computación y Tecnologías de la Información
Universidade da Coruña

anxo.oterod@udc.es, carlos.dafonte@udc.es, diego.fernandez@udc.es, fidel.cacheda@udc.es, manuel.fernandezl@udc.es, francisco.javier.novoa@udc.es

En la actualidad, el uso de redes inalámbricas crece exponencialmente en entornos empresariales de todo tipo. Si bien es cierto que existen una gran cantidad de soluciones en el ámbito de auditoría de redes inalámbricas para grandes organizaciones, las soluciones que existen para las pequeñas empresas son escasas, y esto junto a la falta de conocimientos y experiencia en Tecnologías de la Información (TI) del personal de dichas organizaciones, provoca que este tipo de empresas se encuentren habitualmente en un nivel de riesgo en ciberseguridad alto.

En este contexto desarrollamos una herramienta que tiene como objetivo la auditoría de redes inalámbricas en entornos empresariales, basada en *hardware* de bajo coste y que requiera, únicamente, un nivel básico de conocimientos de TI y ciberseguridad por parte del usuario.

El diseño arquitectónico de la herramienta se basa en un sistema distribuido de dispositivos de bajo coste que permite monitorizar y auditar el entorno inalámbrico y mostrar la información obtenida al usuario de forma inteligible.

En la implementación actual utilizamos Raspberry Pi 3B+ como placas de bajo coste, a las que conectamos antenas Wi-Fi externas, que facilitan la captura de tráfico de red. Posteriormente, procesamos dicho tráfico y los resultados obtenidos se muestran al usuario mediante una interfaz web.

Tras la finalización del desarrollo de la herramienta, hemos realizado pruebas, tanto en un entorno real como en un entorno simulado, lo que nos ha permitido obtener interesantes conclusiones acerca del trabajo realizado.

***Palabras Clave*- Wi-Fi, auditoría, placas de bajo coste**

I. Introducción

En la actualidad, el uso de las redes inalámbricas crece exponencialmente en entornos empresariales de todo tipo, fundamentalmente en las empresas de pequeño tamaño (menos de 10 trabajadores), denominadas microPYMEs. Las características que proporciona esta tecnología (movilidad, facilidad de despliegue y ahorro de coste frente a las redes cableadas) provocan que sean la opción de conectividad preferida en este tipo de organizaciones.

Sin embargo, en este tipo de entidades rara vez se realiza una auditoría o evaluación de seguridad y es habitual que sus miembros no sean conscientes de los riesgos que implica el uso de las redes inalámbricas [1].

Tal y como se describe en "*Survey on WiFi intrastructure attack*" [2], los ataques *"Rogue AP"* o *"Evil Twin"*, *"ARP spoofing"* y *"WiFi MiTM"* son cada vez más habituales. Además, es necesario tener en cuenta las amenazas *DoS* que se implementan a través de ataques de deautenticación, desasociación, falsa autenticación o *"beacon flood"*, son difícilmente evitables.

Los grandes proveedores de infraestructura LAN inalámbrica (WLAN) como Cisco Aironet, HPE-Aruba, Huawei o Fortinet proporcionan soluciones de seguridad avanzadas que ayudan a mitigar estas amenazas y que facilitan la realización de auditorías detalladas. Existen también productos orientados a empresas de tamaño intermedio como Cisco Meraki o Unify de Ubiquiti que también proporcionan productos de monitorización y evaluación de seguridad. Sin embargo, estas soluciones no son habituales en las microPYMEs debido a su coste y al desconocimiento acerca de los riesgos que entraña el uso de WLAN. En España, el 94,8% de las empresas son de este tipo [2] lo que implica que existe una gran cantidad de organizaciones que utilizan tecnologías Wi-Fi en situación de riesgo de seguridad.

En este último año, debido a la pandemia ocasionada por el SARS-CoV-2 (COVID-19), la mayor parte de la población mundial se ha visto obligada a confinarse en sus casas propiciando el uso de redes inalámbricas domésticas para llevar a cabo tareas de trabajo en remoto [3]. Normalmente, estas WLAN son proporcionadas por los proveedores de servicios de Internet y los usuarios desconocen cómo configurarlas o personalizarlas, así como el nivel de seguridad que ofrecen. De nuevo, esta realidad genera una situación de riesgo difícilmente mitigable con soluciones de bajo coste.

En este contexto surge la idea de desarrollar una herramienta de auditoría Wi-Fi, económica y de fácil uso.

II. OBJETIVOS

El objetivo principal del trabajo que estamos presentando es desarrollar un sistema basado en placas de bajo coste que realice auditorías de seguridad inalámbricas con una intervención mínima del usuario.

Para ello establecemos unos objetivos y funcionalidades básicas de la herramienta:

- Extracción de características de cada red detectada: para cada red inalámbrica se muestra una lista de características básicas, así como un estudio de la seguridad de la red y de la calidad y potencia de su señal.
- Inventario de dispositivos: generación de un listado de dispositivos en cada WLAN, previa autenticación, tratando de identificar sus características.
- Interfaz web: implementación de una interfaz web ligera y sencilla que muestre al usuario de forma inteligible el estado de su red desde el punto de vista de seguridad
- Envío de información a un servidor central: desarrollo de un mecanismo de envío de información a un servidor central por medio de una API REST que permita centralizar la información de las diferentes áreas del entorno.

III. DESARROLLO

En este apartado detallaremos el proceso de desarrollo de la herramienta. Inicialmente, presentaremos su arquitectura, las tecnologías utilizadas y, a continuación, explicaremos cada uno de los módulos que la conforman.

A. *Arquitectura*

El sistema está diseñado para ser escalable, es decir, para adaptarse a redes inalámbricas que dan cobertura a superficies de diferente tamaño. En el caso de entornos con una superficie limitada, es suficiente utilizar un dispositivo de auditoría, pero sin embargo si la superficie es extensa, es preciso analizar redes inalámbricas en diferentes localizaciones físicas (diferentes sucursales) o se trata de un edificio de varias plantas, es necesario desplegar varios dispositivos de monitorización o agentes.

Por lo tanto, la arquitectura consta de uno o más dispositivos, llamados agentes, cuya función es monitorizar el tráfico de red, extrayendo información relevante que se envía a un servidor central, tal y como se puede observar en la Figura 1.

Estos agentes envían la información clave recopilada y preprocesada a un servidor de almacenamiento central, utilizando una interfaz API REST. Este repositorio central permitirá, en un futuro, aplicar técnicas de Inteligencia Artificial para detectar anomalías o posibles intrusiones.

La ventaja que proporciona esta aproximación es que el sistema de auditoría se adapta a los cambios y el posible crecimiento de la red.

B. *Tecnologías utilizadas*

En este apartado enumeramos las tecnologías utilizadas para el desarrollo de esta herramienta.

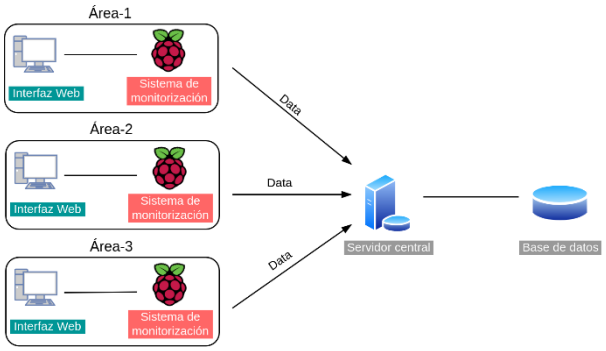


Figura 1. Arquitectura básica

El lenguaje de programación utilizado ha sido *Python*. Su elección se ha fundamentado en la existencia de librerías que han facilitado enormemente el desarrollo, así como por su versatilidad y facilidad para programar de forma ágil y rápida.

Merece una especial mención la librería *Scapy* [5], que permite la manipulación completa de paquetes de red. Es capaz de crear y decodificar paquetes de una gran cantidad de protocolos, transmitirlos, capturarlos o emparejar solicitudes y respuestas.

Proporciona funcionalidades para implementar la mayoría de las tareas clásicas de reconocimiento como *scanning*, *tracerouting*, *probing*, *network discovery*, entre otras.

Para la interfaz web se ha elegido Django [6] y se ha seguido el patrón de diseño *Model-View-Controller*. Para la ejecución de tareas en *background* durante largos periodos de tiempo, sin bloquear la comunicación con el usuario, se ha utilizado Celery [7], que es una librería de código abierto que permite la implementación de colas de tareas de forma asíncrona y facilita la ejecución de operaciones en tiempo real.

En la implementación de la interfaz web se ha utilizado también HTML, CSS, Javascript, JQuery y las librerías Javascript Vis.js [8], para la visualización dinámica de las redes y sus componentes, y Chart.js [9], para la visualización de datos de forma gráfica.

Para el almacenamiento de los datos de la aplicación web se ha elegido PostgreSQL, dada su alto nivel de integración con Django y para el almacenamiento de la información del tráfico recopilado se ha elegido un sistema de gestión de base de datos no relacional como MongoDB, que facilita la lectura y escritura de datos en formato JSON.

Para facilitar el despliegue de la herramienta en múltiples dispositivos hemos empaquetado la aplicación en *docker*. En la implementación actual del sistema utilizamos como plataforma hardware para los agentes una placa *Raspberry Pi 3B+* con un adaptador de red inalámbrico *CSL-AC1200-USB 3.0 Dual Band WLAN Stick"*, que contiene un chip *Realtek8812AU.*

C. *Agentes o dispositivos de captura*

Cada agente se compone de cuatro elementos: un sistema de escucha, un procesador de información, una base de datos y una aplicación web.

El sistema de escucha permite al dispositivo capturar todo el tráfico de red del área en la que se encuentra, de forma pasiva, mediante un adaptador de red externo. Existe

una gran cantidad de tráfico que es irrelevante para propósitos de auditoría y monitorización, por lo que el agente dispone de una funcionalidad de procesado que se encarga de analizar las tramas útiles y extraer la información más importante de cada una de ellas.

Además, cada dispositivo dispone de una base de datos en la que se almacena la información extraída en la escucha, una vez que ha sido procesada.

Cada agente dispone también de una interfaz web, que muestra la información recopilada, procesada y almacenada al usuario mediante un *dashboard* ligero y amigable.

Por último, periódicamente, el agente hace un volcado de la información recopilada al servidor central.

En resumen, cada dispositivo de captura tiene como función procesar y transformar toda la información recibida, guardarla en una base de datos y mostrarla mediante una interfaz web. Además, se encarga de enviar los datos almacenados cada cierto tiempo a un servidor central, de manera que podremos tener toda la información de diferentes áreas de una organización en un sistema centralizado. Este comportamiento permite tener una visión global de todo el entorno.

A continuación, ofrecemos los detalles de implementación de cada uno de estos cuatro módulos funcionales que incluye cada agente.

D. *Módulo de captura de tráfico*

Como hemos comentado previamente, cada agente implementa un sistema de captura de tráfico de red mediante un adaptador de red externo, que proporciona las capacidades físicas para recibir el tráfico inalámbrico. Hemos utilizado la función *sniff()* de la librería *Scapy*, para capturar, de forma pasiva, todo el tráfico de red que está a nuestro alcance y la función *channel_hooper()* nos permite hacer un barrido por un conjunto de frecuencias predeterminadas, de modo que aplicación podrá recibir paquetes de señales inalámbricas de cualquier frecuencia predefinida.

E. *Módulo de análisis de tráfico*

Una vez capturado el tráfico, es necesario analizarlo para clasificarlo y utilizarlo para alguna de las funcionalidades que se especifican a continuación:

- Detección de WLAN y extracción de su información básica: en este caso es necesario procesar las tramas baliza (*beacon frames*) capturadas. Su contenido se transmite en texto claro y contienen información básica acerca de ellas. La función de Scapy *beacon_packet()* permite extraer las características de este tipo de tramas. Entre otros, los datos que se pueden obtener son: SSID, BSSID, fabricante y ratios de transferencia soportados. Además, podemos obtener las características de la señal física que emite un punto de acceso concreto. Para ello, hacemos uso de la cabecera *Radiotap,* que contiene toda la información relacionada con la señal emitida. Entre otros permite obtener los siguientes datos: RSSI, canal utilizado, canal central, frecuencia, ancho de banda utilizado (20 o 40 MHz.), espectro o atenuación de la señal (FSPL, *Free Space Path Loss*). Finalmente, podemos obtener también las características de seguridad de las redes detectadas, por ejemplo: protocolo de seguridad, sistema de autenticación y algoritmo de cifrado.
- Detección de dispositivos conectados a la red (sin autenticación previa): analizando las tramas de datos y control capturadas previamente podemos generar el inventario de dispositivos conectados, a partir de las direcciones MAC origen y destino.
- Detección de tráfico potencialmente peligroso: clasificamos en este caso las tramas de deautenticación debido a que se utilizan en múltiples ataques de DoS. Almacenamos información sobre el punto de acceso al que se han enviado las tramas y cuál es la razón indicada para la deautenticación. Esta información nos ayuda a determinar si las tramas son legítimas o forman parte de un ataque.

F. *Módulo de inventario tras autenticación*

Si bien anteriormente hemos indicado que es posible realizar un inventario de dispositivos sin completar la asociación con un punto de acceso, la información que podemos obtener de estos dispositivos es muy limitada. Además, en ese contexto (no asociados a un punto de acceso) se puede producir un escenario de "nodo oculto", ampliamente referenciado en la literatura, y habría dispositivos que no serían detectados.

Dado que esta herramienta está diseñada como un elemento de auditoría que se instalará en entornos controlados, existe la posibilidad de que el agente se asocie al punto de acceso, disponiendo entonces de conectividad a nivel de enlace en la WLAN a evaluar.

Las funcionalidades que proporciona este módulo son:

- Obtención del inventario de dispositivos, implementando una función que realiza peticiones *arp* a todas las IPs de la red asociada a la WLAN.
- Determinación del sistema operativo de cada dispositivo detectado, mediante las funciones proporcionadas por la librería "nmap3" para Python.

II. Resultados

Debido a que este trabajo está todavía en desarrollo, los resultados obtenidos son bastante limitados. Hemos probado una implementación de la herramienta en la que utilizamos un único agente que hemos probado en dos entornos: un entorno simulado y un laboratorio de investigación de la Facultad de Informática de la Universidade da Coruña (UDC).

A continuación, se detallan los datos más relevantes de ambas auditorías.

A. Entorno simulado

El objetivo de esta prueba es validar la capacidad de la herramienta para realizar todas las funcionalidades indicadas previamente, en especial, la detección de posibles amenazas. Para ello, simulamos un entorno de red inalámbrica de una microPYME, basada en una WLAN proporcionada por un ISP, y lanzamos diferentes simulaciones de ataque.

Los resultados obtenidos son los siguientes:

- Detección de redes inalámbricas: realizamos un escaneo durante 1 hora y detectamos 119 puntos de acceso.
- Simulamos un ataque de tipo *beacon flood* y la herramienta es capaz de detectarlo, al identificar posibles WLAN con SSID sumamente inusuales, sin ningún tipo de seguridad y con valores de señal emitidos completamente anómalos.
- En cuanto a la información referente a las redes inalámbricas asociadas a los puntos de acceso detectados, cabe destacar una gran variabilidad de las características físicas (e.g. RSSI), lo que se corresponde con las previsiones que teníamos. Dichos valores dependen, fundamentalmente, de la distancia a la que se encuentran los puntos de acceso. Los dispositivos finales detectados son, en su mayoría PCs y dispositivos móviles.
- Con respecto a la seguridad de las redes detectadas, aunque existe una cierta variabilidad, en la mayor parte de los casos se trata de redes que implementan el *framework* de seguridad WPA2, con autenticación basada en clave precompartida y un algoritmo de cifrado CCMP.
- En cuanto al análisis de la red inalámbrica propia, nos llama la atención la detección de 10.192 tramas de deautenticación recibidas en el espacio de una hora, lo que claramente indica que un elemento externo está lanzando un ataque DoS de deautenticación. Somos capaces de identificar la dirección MAC del agresor.

B. Entorno de laboratorio en la Facultad de Informática

En este entorno repetimos las tareas de auditoría realizadas anteriormente en el entorno simulado. Los resultados que hemos obtenido son:

- Detección de 42 puntos de acceso y 21 dispositivos finales. La mayor parte de los PAs detectados están vinculados a las redes inalámbricas oficiales de la UDC, como “eduroam”, “udcportal”, “udcodcencia” y “udceventos”.
- Con respecto a la seguridad, se observa que la mayor parte de las redes disponen de una configuración de seguridad fiable, puesto que implementan WPA2 y autenticación basada en usuario. Merece especial atención la red “udcportal”, puesto que podemos observar que se trata de una red abierta, que no implementa ningún tipo de configuración de seguridad, a nivel WLAN (sabemos que se trata de una red abierta que proporciona la UDC para permitir la conectividad a usuarios con problemas en las otras WLAN y que se trata de una solución basada en portal cautivo)

III. Conclusiones

En estos momentos podemos concluir que los objetivos establecidos cuando comenzamos este trabajo se han cumplido. Hemos desarrollado una herramienta con un coste muy contenido, que permite llevar a cabo auditorías de seguridad de redes inalámbricas en diferentes entornos. Somos capaces de detectar las redes inalámbricas que nos radian (tanto propias como ajenas) y determinar sus características, lo que puede ayudar a los usuarios a cambiar características físicas de su red (e.g. canales) y mejorar el rendimiento de su red.

Somos capaces también de detectar tráfico anómalo e inventariar, de forma detallada, los dispositivos de nuestra red inalámbrica.

En un futuro inmediato nuestro objetivo es continuar trabajando en la realización de pruebas y a medio plazo pretendemos integrar este trabajo con otros proyectos centrados en la detección temprana de ataques como [10].

Agradecimientos

Esta investigación ha sido financiada por el Ministerio de Economía y Competitividad de España y fondos FEDER de la UE (Proyecto PID2019-525 111388GB-I00) y por el Centro de Investigación de Galicia "CITIC", financiado por la Xunta de Galicia y la UE (Fondo de Desarrollo Regional Europeo - Programa Galicia 2014-2020), mediante la concesión de ED431G 2019/01.

Referencias

[1] C. Boletsis, R. Halvorsrud, J. B. Pickering, S. Phillips y M. Surridge, «Cybersecurity for SMEs: Introducing the Human Element into Socio-technical Cybersecurity Risk Assesment,» de *16th International Joint Conference on Computer Vision, Imaging and Computre Graphics Theory and Applications*, 2021.

[2] R. Guo, «Survey on WiFi infrastructure attacks,» *International Journal of Wireless and Mobile Computing,* vol. 16, nº 2, 2019.

[3] Organización para la Cooperación y el Desarrollo Económico (OCDE), «Entrepreneurship at a glance,» 2017.

[4] T. Weil y S. Murugesan, «IT Risk and Resilience—Cybersecurity Response to COVID-19,» *IT Professional,* vol. 22, nº 3, pp. 4-10, 2020.

[5] P. Biondi, «Scapy,» 2021. [En línea]. Available: https://www.scapy.net. [Último acceso: 01 06 2021].

[6] A. Makarudze, A. Basset, C. Kirby, W. Vincent, K. Nakamura, M. Eti-mfon y Z. Anderle, «Django Software Foundation,» 2021. [En línea]. Available: https://djangoproject.com. [Último acceso: 01 06 2021].

[7] A. Solem, «Celery - Distributed Task Queue,» 2018. [En línea]. Available: https://docs.celeryproject.org/. [Último acceso: 01 06 2021].

[8] Vis.js Community, «vis.js,» 2021. [En línea]. Available: https://visjs.org. [Último acceso: 01 06 2021].

[9] Chart.js Community, «Chart.js,» 2021. [En línea]. Available: https://chartjs.org. [Último acceso: 01 06 2021].

[10] M. López-Vizcaíno, F. J. Novoa, D. Fernández, V. Carneiro y F. Cacheda, «Early Intrusion Detection for OS Scan Attacks,» de *2019 IEEE 18th International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, 2019.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Técnicas de optimización de redes Wi-Fi centradas en el cliente

Luis Cruz-Piris[a], Jose Manuel Gimenez-Guzman[a], Ivan Marsa-Maestre[a], Susel Fernandez[a], Marino Tejedor-Romero[b]
[a]Departamento de Automática, [b]Departamento de Física y Matemáticas
Universidad de Alcalá
Escuela Politécnica Superior, Campus Universitario, 28805 Alcalá de Henares (Madrid), Spain.
{luis.cruz,josem.gimenez,ivan.marsa,susel.fernandez,marino.tejedor}@uah.es

Aunque las redes de comunicación tienen como objetivo principal dotar de conectividad y proporcionar servicios a los usuarios finales, el papel que tienen estos en su diseño y configuración es testimonial. En la capa del núcleo y distribución de las redes parece razonable que sea así. Sin embargo, en la capa de acceso y, especialmente en redes de acceso inalámbrico donde las acciones de los usuarios alteran de forma sustancial la propia red, no lo es tanto. La hipótesis principal de este trabajo es que el diseño, gestión y configuración de redes de acceso inalámbrico centradas en los usuarios puede proporcionar mejores soluciones respecto a los esquemas actuales. En concreto se propondrá un algoritmo genético para la asignación de canales óptima entre puntos de acceso. Posteriormente, se aplicarán técnicas centradas en los clientes como son la reasignación de clientes a diferentes puntos de acceso y la desconexión temporal de clientes concretos. Los experimentos muestran que dichas técnicas permiten mejorar sensiblemente el rendimiento.

***Palabras Clave*—Wi-Fi, optimización, asignación de canales, algoritmo genético, usuarios**

I. INTRODUCCIÓN

En las dos últimas décadas, las redes de acceso locales inalámbricas han revolucionado la manera en la que los usuarios se conectan a Internet. Entre estas redes, destacan las basadas en la familia de protocolos IEEE 802.11, comercialmente conocidas como redes Wi-Fi. Aunque las redes Wi-Fi disponen de varios modos de funcionamiento, el más popular es el modo infraestructura. En dicho modo, las redes constan de dos tipos de dispositivos de red: puntos de acceso (AP) y estaciones (STA). Ejemplos de STA son ordenadores, teléfonos, tabletas, etc. En el modo infraestructura, cada STA se asocia a un determinado AP y todas sus comunicaciones con el resto de la red se realizan a través de dicho AP. Por ese motivo, comúnmente se indica que las STA son *clientes* de los APs.

La cada vez mayor demanda de servicios de red y coexistencia de redes inalámbricas y usuarios que comparten una misma banda de frecuencias del espectro radioeléctrico ha provocado que el diseño y operación de las redes Wi-Fi sea un problema al que la comunidad científica deba atender. Así pues, la casi totalidad de los trabajos relacionados con las redes Wi-Fi ponen el foco en los APs. Sin embargo, este trabajo parte de la hipótesis de que, si los clientes finales son las STAs, estas deben tener un papel más relevante en el funcionamiento de la red. Es habitual que se asuma que las STAs se asocien al AP del cual reciben una mayor potencia de señal recibida, siendo pocos los trabajos que no hacen dicha suposición. Por ejemplo, en [1] se tiene en cuenta la asociación STA-AP de manera conjunta con la asignación del canal en el que opera cada punto de acceso. De hecho, la asignación o selección del canal en el que opera cada AP ha sido profusamente estudiada, por la complejidad del problema y el impacto que tiene en las prestaciones de la red. En [2] se recogen los principales esfuerzo realizados en este campo, clasificando las técnicas en dos categorías: centralizadas y no coordinadas. No obstante, desde la publicación de dicho estado del arte, han aparecido numerosas aportaciones a este problema [3], [4], [5], [6]. Pese a los numerosos estudios, la asignación de canales en redes Wi-Fi no es un problema resuelto dada su complejidad. Una de las fuentes de esa complejidad en la asignación de canales Wi-Fi es el hecho de que los canales adyacentes en los que puede operar un AP se solapan parcialmente entre sí.

Este trabajo se centra en la asignación de canales en redes Wi-Fi en modo infraestructura dándole un mayor protagonismo al cliente final, siendo sus principales contribuciones las siguientes:

1) Propuesta de un algoritmo genético para conocer las prestaciones máximas que puede ofrecer una red Wi-Fi mediante una selección óptima de canal.
2) Estudio y evaluación de técnicas de asignación de

canales óptima cuando se permite que las STAs no se asocian automáticamente al AP del que reciben una mayor potencia de señal sino al que maximiza el rendimiento de la red.

3) Adicionalmente, también se contempla la opción de realizar una desconexión selectiva de STAs de manera temporal para que así, actuando de manera solidaria temporalmente, otras STAs de la red puedan mejorar su rendimiento.

La estructura del trabajo se indica a continuación. La siguiente sección describe las técnicas de optimización basadas en los clientes, mientras que las secciones III y IV se centran en la descripción de los experimentos y resultados. Finalmente, la sección V concluye el trabajo y define las próximas líneas de estudio.

II. TÉCNICAS DE OPTIMIZACIÓN CENTRADAS EN LOS CLIENTES

Dado que la posibilidad de configurar hasta x parámetros en y nodos de cada escenario inalámbrico es un problema combinatorio de tipo NP-hard [7], las soluciones con un enfoque centralizado y basada en metaheurísticas son muy frecuentes. El uso de algoritmos evolutivos, especialmente genéticos, en este contexto no es nuevo. En [8] los autores buscan ubicaciones óptimas para desplegar puntos de acceso con el objetivo de mantener unos niveles mínimos de señal en todo la zona.

En este trabajo se ha realizado un enfoque diferente partiendo de ubicaciones ya fijadas y variando la asignación de canales que utilizada cada punto de acceso. En la sección A se muestra el algoritmo genético diseñado con este fin. Posteriormente, en las secciones B y C se explicarán, de forma resumida, dos técnicas para variar las configuraciones de la red inalámbrica centradas en los clientes.

A. Asignación de canales basada en un algoritmo genético

En una red Wi-Fi tenemos un conjunto de n puntos de acceso (AP_i) a los que se les debe asignar un canal (ch_j) de entre los m disponibles en el conjunto CH. Para el caso de la banda de 2.4 GHz, $m = 11$. Teniendo en cuenta esto, se define el individuo del algoritmo genético como la sucesión de canales asignados a cada APs, es decir, $I = [ch_1, ch_2, \ldots, ch_n]$ donde $ch_x \in CH$ y cada gen se corresponde con un AP_i. Por tanto, la longitud del individuo será n (número de APs de la red).

Los operadores definidos para este GA se muestras a continuación:

- Generación de la población inicial: Para cada gen que compone cada individuo se le asigna de forma aleatorio un valor del conjunto CH.
- Operador de cruce: Dados dos individuos (padre 1 y padre 2), se genera un vector de la misma longitud de los individuos formado por unos y ceros. El hijo es el resultado de tomar los valores de las posiciones donde este vector vale cero del padre 1 y en las posiciones que tiene el valor uno del padre 2.
- Operador de mutación: Se evalúa cada gen del individuo y con una probabilidad P_m se asigna a ese gen un nuevo valor del conjunto CH.
- Función de evaluación: Representa la utilidad total de la red. Se detalla en la sección III.A.

B. Reasignación de clientes (CR)

En el modo de funcionamiento pasivo de la redes Wi-Fi, las STA (clientes) escuchan esperando la llegada de balizas generadas por los APs. En estas balizas, cada AP proporciona información relativa a sus modos de funcionamiento e identificación. En el caso de que una STA tenga varias alternativas para conectarse a una misma red, de forma generalizada, inicia el proceso de conexión al AP que le ofrece la señal con mayor potencia.

Partiendo de la hipótesis de que para redes muy densas este tipo de asignación puede no ser la mejor opción, se diseña una metodología de reasignación de cliente a AP que sigue los siguientes pasos:

1) Desde el punto de vista de cada STA se establecen los APs que están en su radio de cobertura.
2) Cada STA evalúa la calidad de su conexión en base al parámetro RSSI (*Received Signal Strength Indicator*).
3) Se establece un valor umbral con la calidad mínima deseable para cada STA.
4) Cada STA que tiene tiene una conexión con un valor de RSSI inferior al umbral, se desconecta del AP actual y establece una conexión con el siguiente AP más próximo (con mejor valor de potencia).
5) Se evalúa el nuevo valor de utilidad de ese nodo. Si es mejor, la conexión se hace permanente, si empeora, el algoritmo prueba con el siguiente AP.

C. Desconexión selectiva de clientes (SD)

En situaciones donde el número de STA es muy numeroso y hay un alto grado de interferencias, el valor de RSSI de cada STA suele ser bajo. Esto afecta directamente a la calidad de la conexión, llegando incluso a la degradación del servicio sea tal que pudiera ser equivalente a que una STA no estuviera asociado a su AP. De esta situación se plantea la hipótesis de que los clientes puedan elegir periodos de tiempo de inactividad, donde evitarían realizar cualquier tipo de emisión y, por tanto, dejarían de actuar como señales interferentes sobre el resto. Si además, esta desconexión selectiva se realiza por grupos, la utilidad total de la red podría aumentar previsiblemente.

Con este fin se diseña el modo de funcionamiento SD. Cada STA que tiene un valor RSSI inferior a un determinado valor, puede agruparse con otras STA en la misma situación. Estos grupos de STAs deciden intervalos de tiempo en los que pasaran a un modo inactivo. Este comportamiento se replica entre grupos de STA coordinadas. En esta fase de la investigación, donde se persigue validar la funcionalidad de este sistema, la gestión de las STA se realiza de forma centralizada.

III. EXPERIMENTOS

Para aplicar las primeras pruebas sobre las técnicas propuestas se han tenido que desarrollar dos tipos de elementos: un modelo para la propagación de señales e interferencias, y un conjunto de escenarios representativos sobre los que realizar los experimentos. En esta sección se mostrarán, de forma resumida, las propuestas realizadas en estos sentidos.

A. Modelo de propagación e interferencias

Las redes de comunicaciones son modeladas habitualmente como grafos. Para este problema en concreto se ha optado por utilizar dos grafos: el grafo de conectividad (G) y el grafo de interferencias (I). G representa la asociación entre los clientes (STA) y los puntos de acceso (AP) de la red Wi-Fi basada en infraestructura.

Como las interferencias tendrán diferentes intensidades, I es un grafo ponderado donde el coste de cada enlace representa ese valor. Dados dos nodos x e y, funcionando en los canales i y j, respectivamente (i y $j \in \{1, \ldots, 11\}$, puede ser definida como:

$$I(x,y) = P_t + G_t + G_r - L - P_{loss} + \psi + W(i,j), \quad (1)$$

donde P_t es la potencia de transmisión en dBm, G_t y G_r representan la ganancias de las antenas de emisión y recepción en dB, L representa las pérdidas de potencia debida a los posibles obstáculos en dB y P_{loss} representa las pérdidas de propagación en dB (basado en la distancia y alturas de las antenas [9]). Ya que las estaciones y los puntos de acceso no están transmitiendo todo el tiempo, se estable el parámetro ψ como el índice de actividad. Por último, $W(i,j)$ representa la interferencia entre los canales i y j debido a su solapamiento en frecuencia. Los valores utilizados para definir $W(i,j)$ son los obtenidos en base a experimentos en [10]. Ya que el uso de redes basadas en frecuencias de 2.4 GHz son las más habituales, y las que por su radio de cobertura generan más conflictos, será esta banda en la que se centre el modelo realizado.

A partir de los grafos G e I es sencillo obtener el valor de $SINR_i$ (*signal-to-interference-plus-noise ratio*) como el cociente entre la señal deseada y la suma del resto de señales interferentes que recibe el nodo i. La utilidad de cada nodo (U_i) es un valor entre 0 y 1, obtenido a partir de su $SINR_i$ tal y como se muestra en [11]. La función objetivo de este modelo calcula el valor de utilidad de cada nodo y devuelve la suma de utilidades tal que $U = \sum_{\forall i} U_i$.

B. Escenarios de pruebas

Las pruebas realizadas se han llevado a cabo utilizando la distribución de APs (26 en total) existente en la primera planta de la Escuela Politécnica Superior de la Universidad de Alcalá. Se han generado un conjunto de 12 escenarios diferentes, variando el número de estudiantes y sus posiciones (teniendo en cuenta si se encuentran dentro o fuera de las aulas). Estos escenarios pueden verse en detalle en trabajos anteriores como [3], [5], [7], [6].

En la fig.1 se muestran 3 de los 12 escenarios donde se pueden apreciar como los APs (puntos azules) siempre están ubicados en las mismas posiciones, mientras los clientes (resto de puntos) varían de posición y aumentan según el escenario.

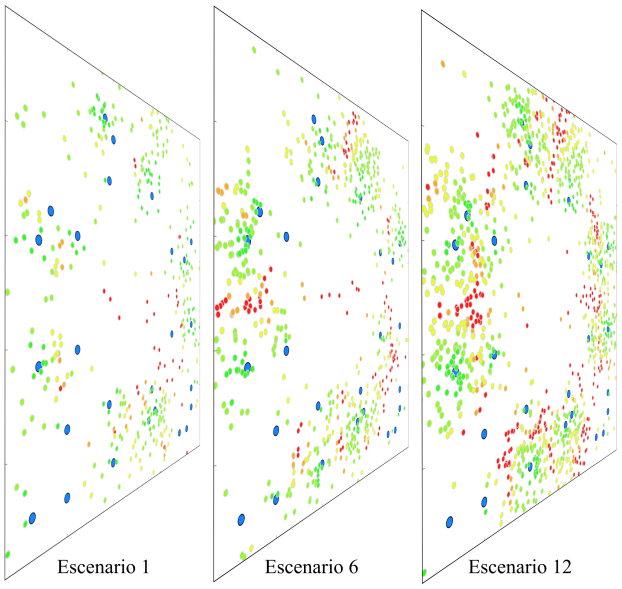


Fig. 1. Representación de tres escenarios donde los círculos azules representan los puntos de acceso y el resto de círculos los clientes (coloreados según su valor de utilidad).

C. Pruebas realizadas

La asignación de canales a cada punto de acceso ha sido abordada en anteriores trabajos utilizando optimizadores basado en LCCS (*Least Congested Channel search*), SA (*Simulated Annealing*)[3] y CRO-SL (*Coral Reefs Optimization with Substrate Layer*)[6]. En este trabajo se introduce los algoritmos genéticos (GA) como técnica de optimización base, además de aplicar sobre los mejores resultados obtenidos, los métodos de reasignación de clientes (CR) y desconexión selectiva (SD).

En un primer conjunto de pruebas se realiza la optimización de los canales Wi-Fi para los mismos 12 escenarios de anteriores trabajos utilizando el algoritmo genético propuesto en la sección A. Para cada uno de los escenarios realiza la elección de canales utilizando el GA una probabilidad de mutación del 2%. Estos resultados se compararán con los obtenidos por otros métodos.

Un segundo bloque de pruebas se centra en, a partir de la mejor configuración de canales obtenidas para con CRO-SL y GA, se aplicar de forma centraliza una resignación de clientes y una desconexión selectiva, evaluando el valor de la función de utilidad global de cada escenario.

IV. RESULTADOS INICIALES

La tabla I compara la propuesta de utilizar un algoritmo genético para la optimización de la asignación de canales Wi-Fi en escenarios complejos con los resultados obtenidos en trabajos anteriores [6]. Los resultados de la propuesta mejoran siempre los obtenidos con LCCS y SA. Además, se aproximan e incluso mejoran en determinados

escenarios a los valores obtenidos con CRO, siendo GA una técnica computacionalmente menos costosa.

Tabla I
VALORES DE UTILIDAD MÁXIMA OBTENIDOS PARA MÉTODO DE OPTIMIZACIÓN DE ASIGNACIÓN DE CANALES WI-FI UTILIZADOS.

Escenario	LCCS	SA	CRO-SL	GA
1	214,05	244,68	**251,17**	245,73
2	206,91	234,40	**241,29**	236,98
3	233,19	239,09	**249,69**	243,43
4	285,93	356,00	**365,24**	358,44
5	314,16	356,68	**377,13**	371,56
6	315,63	385,10	375,96	**378,26**
7	378,83	494,36	**512,93**	502,39
8	396,42	543,48	556,20	**557,11**
9	371,06	501,92	507,32	**510,07**
10	468,93	605,03	**598,94**	593,04
11	472,61	601,77	**636,32**	613,75
12	481,65	586,66	**616,83**	600,61

Para analizar las posibles ventajas de los métodos de reasignación de clientes (CR) y desconexión selectiva (SD) se han tomado como base las configuraciones obtenidas con los optimizadores CRO-SL y GA en el paso anterior, ya que son los que mejor valor de utilidad han ofrecido. A partir de estos escenarios en lo que se había ya estábamos en valores máximos, se ha procedido a analizar la repercusión que tendrían los métodos CR y SD. Los resultados obtenidos para todos los escenarios de estudio se muestran en la tabla II.

Tabla II
MEJORAS OBTENIDAS CON LOS MÉTODOS DE REASIGNACIÓN DE CLIENTES (CR) Y DESCONEXIÓN SELECTIVA (SD).

Escenario	CRO-SL			GA		
	OP	CR	SD	OP	CR	SD
1	251,17	252,43	267,62	245,73	248,17	260,55
2	241,29	241,31	253,48	236,98	237,06	252,74
3	249,69	250,30	259,56	243,43	243,43	250,75
4	365,24	365,24	392,91	358,44	361,33	380,99
5	377,13	378,39	409,72	371,56	373,73	406,41
6	375,96	377,06	414,58	378,26	379,58	415,28
7	512,93	514,14	564,95	502,39	502,60	551,62
8	556,20	558,07	599,20	557,11	558,97	600,85
9	507,32	507,99	557,61	510,07	511,61	555,84
10	598,94	607,28	679,18	593,04	594,52	667,32
11	636,32	637,96	694,78	613,75	617,13	684,80
12	616,83	618,27	696,17	600,61	601,89	682,07

Aunque la configuración base de los APs en estos escenarios es un caso extremo en el sentido de que se ha llegado previamente a unos valores de utilidad muy elevados con procesos de optimización intensivos, se puede apreciar como los métodos CR y SD proporcionan mejoras sensibles. Para el caso de la reconexión selectiva se consiguen mejoras de entre el 0,5% y el 1%. Estas mejoras no son demasiado grandes, pero se deben de poner en el contexto de un escenario estático y muy optimizado. Por otro lado, la desconexión selectiva de las STA que tiene peor RSSI del escenario proporciona mejoras en la utilidad global de entorno al 10%, llegando a tener incluso valores máximos del 13,56%.

V. TRABAJOS FUTUROS

Nuestro objetivo es explorar la posibilidad de trasladar el peso de la configuración de una red Wi-Fi a los clientes, frente a los mecanismos tradicionales centrados en la propia red. En este artículo se presentan las primeras hipótesis y resultados de nuestro trabajo en esta línea, quedando de manifiesto que es un desafío importante pero con el potencial de ser muy fructífera. Las técnicas utilizadas de reasignación y desconexión selectiva de clientes, aunque llevadas a cabo desde un punto de vista centralizado, muestran que proporcionan mejoras sensibles. Los siguientes pasos de la investigación se centrarán en la aplicación de estas técnicas de forma distribuida, a partir de la percepción de cada cliente.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por las ayudas SBPLY/19/180501/000171 de la Junta de Comunidades de Castilla-La Mancha y FEDER, y CM/JIN/2019-031 de la Comunidad de Madrid y la Universidad de Alcalá.

REFERENCIAS

[1] B. P. Tewari and S. C. Ghosh, "Joint frequency assignment and association control to maximize the aggregate throughput in ieee 802.11 wlan," *Wireless Personal Communications*, vol. 94, no. 3, pp. 1193–1221, 2017.

[2] S. Chieochan, E. Hossain, and J. Diamond, "Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 1, 2010.

[3] E. de la Hoz, J. Gimenez-Guzman, I. Marsa-Maestre, and D. Orden, "Automated Negotiation for Resource Assignment in Wireless Surveillance Sensor Networks," *Sensors*, vol. 15, no. 11, pp. 29 547–29 568, Nov. 2015.

[4] M. Abusubaih, "Using Partially Overlapping Channels in Home 802.11g WLANs," *Wireless Personal Communications*, vol. 88, no. 2, pp. 295–303, May 2016.

[5] E. De La Hoz, I. Marsa-Maestre, J. M. Gimenez-Guzman, D. Orden, and M. Klein, "Multi-agent nonlinear negotiation for wi-fi channel assignment." in *AAMAS*, 2017, pp. 1035–1043.

[6] C. Camacho-Gómez, I. Marsa-Maestre, J. M. Gimenez-Guzman, and S. Salcedo-Sanz, "A Coral Reefs Optimization algorithm with substrate layer for robust Wi-Fi channel assignment," *Soft Computing*, vol. 23, no. 23, pp. 12 621–12 640, Dec. 2019.

[7] D. Orden, J. Gimenez-Guzman, I. Marsa-Maestre, and E. de la Hoz, "Spectrum Graph Coloring and Applications to Wi-Fi Channel Assignment," *Symmetry*, vol. 10, no. 3, p. 65, Mar. 2018.

[8] L. Nagy and L. Farkas, "Indoor base station location optimization using genetic algorithms," in *11th IEEE International Symposium on Personal Indoor and Mobile Radio Communications. PIMRC 2000. Proceedings (Cat. No.00TH8525)*, vol. 2, 2000, pp. 843–846 vol.2.

[9] D. Green and A. Obaidat, "An accurate line of sight propagation performance model for ad-hoc 802.11 wireless lan (wlan) devices," in *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002 (Cat. No.02CH37333)*, vol. 5, 2002, pp. 3424–3428 vol.5.

[10] S. W. Ng and T. Szymanski, "Interference measurements in an 802.11n wireless mesh network testbed," in *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2012, pp. 1–6.

[11] A. Bazzi, "On uncoordinated multi user multi rat combining," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1–6.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Automatización del cálculo del nivel de seguridad de un entorno IoT basado en el inventario y las vulnerabilidades intrínsecas del sistema

Marc Salinero, Julia Sánchez, Guiomar Corral.
Departamento de Ingeniería, Grupo de Investigación en Internet Technologies and Storage (GRITS),
La Salle Campus Barcelona – Universidad Ramon Llull (URL)
Quatre Camins 30, 08022, Barcelona.
marc.salinero@students.salle.url.edu, j.sanchez@salle.url.edu, guiomar.corral@salle.url.edu.

Este documento presenta la implementación de un bloque encargado de calcular el nivel de seguridad asociado a un entorno IoT basándose en los dispositivos presentes y en las vulnerabilidades intrínsecas del mismo. Este bloque forma parte de un sistema más complejo encargado de calcular el riesgo total del entorno IoT añadiendo el resultado de: (1) la realización de ciertos tests contra los dispositivos encontrados y, (2) la correlación de los logs almacenados fruto de la monitorización continua de dicho entorno.

***Palabras Clave*- jitel, telemática, IoT, ciberseguridad, riesgos**

I. Introducción

El paradigma IoT (*Internet of Things*) ha ganado importancia en los últimos años debido a las ventajas que proporciona al mejorar nuestra calidad de vida. Los dispositivos IoT obtienen información del entorno que, tratada adecuadamente, proporciona cierto grado de inteligencia permitiendo tomar decisiones difíciles de manera más fácil y, por tanto, ejecutar tareas diarias con la mínima intervención humana. IoT convierte los hogares, edificios, hospitales, ciudades, industrias, entre otros, en sistemas inteligentes capaces de obtener el conocimiento del entorno y aplicarlo para su adaptación de acuerdo a las necesidades de los habitantes.

El IoT no para de crecer, cada vez más todos los dispositivos que tenemos o tendremos estarán conectados entre ellos y al *cloud*. Se multiplicarán las conexiones, la información transmitida a través de la red y, desafortunadamente, los ataques hacia estos dispositivos. Estos dispositivos pueden transmitir información crítica o confidencial y, por tanto, el nivel de seguridad que deben tener es elevado.

El grupo de investigación en *Internet Technologies and Storage* (GRITS) de La Salle Campus BCN-URL ha dedicado parte de sus esfuerzos, desde hace algún tiempo, a investigar sobre IoT, concretamente sobre ciberseguridad en entornos IoT. La gran dependencia que mantiene la sociedad actual con este tipo de entornos, genera la necesidad de estudiar cómo ofrecer un buen nivel de seguridad al usuario final y de la manera más eficiente posible.

En el grupo de investigación GRITS, se han realizado varias investigaciones y se ha participado en varios proyectos que han aportado información suficiente para plantear un sistema que proteja de manera adecuada un entorno IoT.

- Del proyecto FINESCE [1], se extrajeron las implicaciones de seguridad y contramedidas a aplicar en entornos *cloud*, lugares donde se pueden encontrar aplicaciones IoT.
- De investigaciones realizadas sobre ecosistemas IoT y sistemas ICS (*Industrial Control Systems*), se vieron las tendencias y necesidades de ciberseguridad para este tipo de entornos. En el proyecto SPRINT 4.0 [2] quedaron al descubierto los retos de seguridad, las amenazas y la necesidad de implementar sistemas robustos en el sector de la Industria 4.0, así como la necesidad de aplicar un conjunto de mejores prácticas para un correcto diseño de un sistema seguro.
- De otras investigaciones y proyectos como SmartCampus [3], u otros centrados en hogares inteligentes, se aprendieron las diferencias entre las comunicaciones de entornos IoT en ámbitos

distintos, y por tanto la necesidad de adaptación del sistema y su seguridad según el ámbito IoT.

Toda esta investigación muestra que existen problemas y vulnerabilidades en todos los ámbitos principales de la actualidad, que existe un gran incremento en la conexión de dispositivos de naturaleza heterogénea a Internet (para procesos críticos de la sociedad y la industria), que las tecnologías emergentes suponen un cambio constante y nuevos vectores de ataque, y que los ataques cada vez son más sofisticados y difíciles de detectar. De manera que el proceso de securización de entornos IoT no es nada fácil.

El problema no es sólo la inmensa cantidad de dispositivos conectados a la red, sino que también se debe tener en cuenta que existen dispositivos con sistemas o software de seguridad más avanzado y dispositivos simples, como bombillas o sensores, que no tienen detección de ataques y son más vulnerables. También es importante considerar la seguridad de la red y controlar los accesos, así como la seguridad de los proveedores de servicios que dan acceso a los dispositivos IoT de manera remota y proveedores *cloud,* que albergan las aplicaciones que el entorno IoT pueda utilizar para su gestión y mantenimiento.

Además, el hecho de que haya múltiples fabricantes también dificulta el proceso. Cada fabricante diseña sus productos con sus propios protocolos de seguridad y protocolos de comunicación, lo que hace que sea muy complicado crear un estándar a seguir para aplicar seguridad [4][5]. A pesar de esta dificultad, gobiernos e instituciones, entre ellas Estados Unidos y la fundación OWASP, han comenzado a desarrollar proyectos para concienciar sobre estos riesgos. El proyecto llevado a cabo en Estados Unidos busca concienciar a los ciudadanos proponiendo unos pasos a seguir para prevenir posibles brechas de seguridad en los dispositivos. Por otro lado, la fundación OWASP [6] alerta sobre las 10 vulnerabilidades más comunes en entornos IoT de manera anual, y proporciona información y guías para diseñar, desarrollar configurar y testear dispositivos IoT.

Se pueden encontrar múltiples soluciones a los diferentes retos de seguridad y problemas que plantea un entorno IoT, pero hasta hace pocos meses, ningún estándar aprobado. Actualmente, existen estándares como el europeo ETSI EN 303 645 [7] con el objetivo de proporcionar unas mejores prácticas a aplicar en entornos IoT. A continuación, se listan ejemplos de mejores prácticas o recomendaciones extraídas de [7] y [8] que, a pesar de parecer obvias, son importantes y buscan concienciar y ayudar a asegurar una red contra posibles ataques:

- Utilizar credenciales robustas y cambiarlas con frecuencia; no usar contraseñas por defecto.
- Cifrar los datos transmitidos a través de la red.
- Tener los dispositivos al día de actualizaciones de software/firmware o antivirus.
- Almacenar de forma segura los parámetros críticos de seguridad.
- Asegurar la integridad del software.
- Revisar los permisos de las aplicaciones (como las de los wearables y smartphones).
- Segmentar la red para tener mayor control de lo que sucede y aislar las zonas más sensibles.
- Asegurarse de que los datos personales sean seguros.
- Hacer que el sistema sea resiliente a los cortes.
- Utilizar Firewalls que obstaculicen el tráfico sospechoso y hagan seguimiento de eventos.
- Controlar el acceso a dispositivos.
- Hacer que los usuarios puedan eliminar sus datos fácilmente.
- Hacer que la instalación y mantenimiento de los dispositivos sea sencilla.
- Implementar métodos de análisis de riesgos para estimar la seguridad del entorno IoT.
- Tener medios para generar *reports* de las vulnerabilidades.
- Hacer auditorías de seguridad periódicamente para saber el estado de todos los dispositivos en términos de protección, control y medidas de seguridad.

Aparte del estándar mencionado anteriormente, existe el NISTIR 8259 (del departamento de comercio de los estados unidos) [9] el cual se centra más en dar recomendaciones a los fabricantes para securizar sus dispositivos. El estándar ofrece seis consejos para mejorar la seguridad de los dispositivos, cuatro de ellos se deben realizar antes de lanzar el producto al mercado, y los dos restantes se aplican una vez ya se ha puesto el producto en venta.

El trabajo presentado en este documento utiliza las recomendaciones de "*Implementar métodos de análisis de riesgos*" y "*Hacer auditorías de seguridad periódicamente*" y propone un sistema que calcula el riesgo de seguridad de un entorno IoT de manera automática. En la Sección II, se presenta el sistema y se definen los diferentes bloques funcionales que lo conforman. En la Sección III, se explica la implementación del primero de los bloques funcionales, "*Inventario del sistema y nivel de seguridad intrínseco*", detallando su diseño, los cálculos, la implementación de código y los resultados obtenidos sobre el entorno de testeo. En la Sección IV se muestran las conclusiones del trabajo realizado y, finalmente, la Sección V detalla las líneas futuras que se han observado una vez realizada la implementación del primer bloque funcional del sistema.

II. Cálculo de Riesgos en Entornos IoT

El objetivo del sistema que se propone es proporcionar el cálculo de riesgo de un entorno IoT. El sistema está compuesto por los bloques mostrados en la Fig. 1.

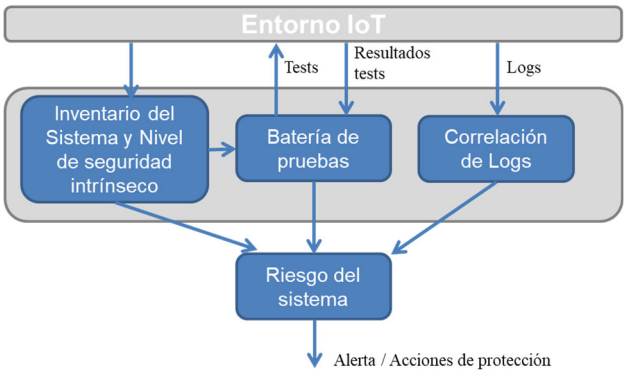


Fig. 1. Sistema para el cálculo de riesgo de un entorno IoT

Inventario del Sistema y Nivel de seguridad intrínseco. Inventario de dispositivos (HW), SW, tecnologías de comunicaciones utilizadas, etc. Proporciona un valor que dará una idea del nivel de seguridad intrínseco del sistema según las vulnerabilidades que tengan los dispositivos. Debe actualizarse automáticamente al añadir un dispositivo para permitir actualizar el cálculo del riesgo total.

Batería de pruebas. Conjunto de tests realizados al entorno IoT para determinar la existencia de ciertas amenazas y poder afinar el cálculo de riesgo. Las vulnerabilidades encontradas en el bloque anterior alimentan este segundo bloque.

Correlación de Logs. De dispositivos IoT, Gateway IoT, PCs, Smartphone/Tablet, dispositivos de red, dispositivos de seguridad de red, etc. Proporciona información del estado del sistema y permite detectar posibles amenazas y comprobarlas.

Riesgo del sistema. Mediante la información recogida de los tres bloques anteriores, se genera un cálculo de riesgo del sistema.

Finalmente, y como se aprecia en la Fig. 1, del bloque "*Riesgo del sistema"* salen Alertas y Acciones de Protección. Esta funcionalidad se propone como implementación a largo plazo con el objetivo de conseguir un sistema de predicción temprana, que sea capaz de aprender del entorno de manera automática y auto-protegerse, generando e instalando automáticamente reglas en los firewalls o ejecutando las acciones necesarias determinadas previamente con todo el proceso de análisis y cálculo del riesgo.

Para poder hacer frente a este gran reto, es necesario ir resolviendo cada uno de los bloques por separado, convirtiendo el sistema en pequeños subsistemas que posteriormente se enlazarán. Y, además, será necesaria la automatización de las tareas para que todo el conjunto sea viable.

Este documento detalla una primera aproximación de la implementación del bloque "*Inventario del Sistema y Nivel de seguridad intrínseco*".

III. Implementación del bloque "Inventario del Sistema y Nivel de Seguridad Intrínseco".

El objetivo de automatizar las tareas que se realizan en un análisis de riesgos es el de mejorar y optimizar estos procesos. Es obvio que este proceso automatizado siempre debe estar supervisado por una persona por si se da el caso de que el resultado no es del todo fiable o algún procedimiento no se ha hecho correctamente. Pero el hecho de que esté automatizado permite un ahorro significativo de los recursos de una empresa, y, además, consigue el resultado final con mucho menos tiempo que si se hiciera manualmente.

La diferenciación que presenta este proyecto respecto a los análisis de riesgos convencionales, aparte de la automatización, es que se realiza sobre un entorno IoT y no sobre una red convencional.

Hay que decir que el procedimiento a seguir es bastante parecido al de un análisis de riesgos convencional, pero se deben tener en cuenta ciertos parámetros y procedimientos que, por el simple hecho de ser un entorno IoT, pueden cambiar considerablemente. Un ejemplo de ello podría ser el número de dispositivos que forman la red, el cual se puede disparar considerablemente en un entorno IoT y, consecuentemente, también se puede disparar el volumen de tráfico que circula por la red. Además, se debe tener en cuenta que muchos de estos dispositivos pueden ser sensores o pequeños microprocesadores, los cuales no tendrán sistemas de seguridad avanzados como los que puede tener un ordenador o los servidores.

A. Formato y estilo

El trabajo se ha dividido en cuatro grandes bloques. Con el fin de automatizar estos cuatro bloques y que funcionen conjuntamente, se ha escrito un script en Python, el cual sólo hay que ejecutar y él solo se encargará de ejecutar los cuatro bloques y mostrar los resultados obtenidos.

El resultado que se quiere conseguir mediante este script es obtener un valor cuantificado de la seguridad de la red. Este valor será mostrado en una escala del 1 al 10, donde 10 es el más crítico, y servirá para hacernos una idea del riesgo existente en nuestra red o entorno IoT. Posteriormente, este valor sería utilizado en el sistema presentado en la Fig. 1 para obtener el valor de riesgo del sistema.

En la Fig. 2, se muestra un diagrama de bloques que representa el funcionamiento del script que implementa el bloque "*Inventario del Sistema y Nivel de Seguridad Intrínseco*".

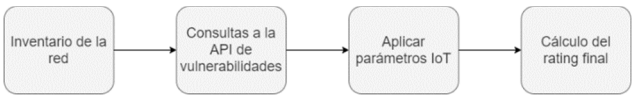


Fig. 2. Diagrama de bloques del script

Inventario de la red. El primer bloque consiste en hacer un escaneo de toda la red, con el fin de saber qué dispositivos están conectados y sus características. Además, a la vez que se descubren los dispositivos, también se escanean las posibles vulnerabilidades que puedan tener.

Consultas a la API de vulnerabilidades. Ahora que ya tenemos la información imprescindible para cada dispositivo, toca buscar más información sobre cada vulnerabilidad, haciendo consultas a una API. Buscamos el vector asociado a cada vulnerabilidad con el que vamos a calcular el riesgo que presenta.

Aplicar parámetros IoT. Este bloque es de los más importantes ya que es el que se encarga de tener en cuenta el hecho de que es una red IoT y no una convencional. Se modifica el vector obtenido en el bloque anterior, y se añaden ciertos parámetros que en una red convencional no se tendrían en cuenta, pero en un entorno IoT es imprescindible contar con ellos.

Cálculo del rating final. Por último, una vez tenemos los vectores nuevos adaptados al entorno IoT, toca hacer los cálculos pertinentes para pasar de vector a un valor comprendido entre el 1 y el 10. De esta forma, después de realizar todos los cálculos para todas las vulnerabilidades, sólo se tendrá que hacer la media de estos valores y se obtendrá el *Score* o *Rating final* del entorno IoT.

B. Inventario de la red

El script implementado está pensado para ejecutarse en un entorno Linux, que contenga el software Nmap [10], ya que el script lo usa. En este trabajo se ha usado Kali Linux que ya lleva incorporadas todas las herramientas necesarias.

Se usa Nmap porqué es una herramienta muy potente cuando se trata de escanear redes, además cuenta con la ventaja de que es Open Source.

Gracias a que permite el uso de scripts, se pueden incrementar en gran medida las capacidades de esta herramienta, y podemos conseguir que, aparte de escanear los puertos abiertos de un dispositivo y averiguar su sistema operativo, con el uso de los scripts podemos llegar a escanear las vulnerabilidades. Para ello se ha hecho uso de un script llamado **nmap-vulners** [11].

Existen otras herramientas más potentes para detectar vulnerabilidades como viene siendo el caso de Nessus [12], pero el inconveniente que presenta es que sólo se puede usar con una interfaz de usuario y no por CLI (*Command Line Interface*), aparte de que es una herramienta de pago.

Entonces, si el objetivo del proyecto es automatizar todo este proceso, una herramienta que se controla por CLI es mucho más fácil de scriptar que una herramienta basada en una GUI (*Graphical User Interface*).

La instrucción utilizada para generar el inventario y encontrar las vulnerabilidades de todos los dispositivos es la siguiente:

```
nmap --script nmap-vulners -sV -O -oX output.xml 10.14.1.0/24
```

Como se puede ver anteriormente, la instrucción hace uso de cuatro modificadores diferentes. El primero es el *--script* que indica al Nmap que se quiere utilizar el script *nmap-vulners*, el cual será el encargado de mostrarnos toda la información relacionada con las vulnerabilidades de cada dispositivo.

El segundo modificador es el *-sV* que se utiliza para determinar la versión y el servicio de cada puerto. Por ejemplo, si el puerto se trata de un Apache o cualquier otro servicio, indicando en la mayoría de los casos su versión.

El tercer modificador es el *-O* que sirve para habilitar el escaneo de Sistema Operativo, es decir, nos dirá si el dispositivo en cuestión se trata de un Windows, un Linux, etc.

Finalmente, el cuarto modificador, el *-oX*, simplemente sirve para facilitar el tratamiento de los datos en el siguiente bloque. Lo que hace es parsear la salida del Nmap en formato XML y guarda esta salida en un fichero específico.

Y para terminar tenemos que introducir la red dónde queremos lanzar el escaneo del Nmap.

Para procesar la información que nos devuelve la instrucción en formato XML, primero se ha cogido el archivo y mediante el uso de una librería que permite convertir el formato XML del fichero a diccionario, se ha convertido el contenido a formato diccionario. Esta librería se llama **xmltodict**. Trabajar con información en formato diccionario en Python es mucho más sencillo que tratar de leer y acceder a los datos de un archivo XML.

La salida del Nmap en formato XML ha dificultado el tratamiento de la información ya que, dependiendo del host no siempre se accede de la misma manera al CVE de una vulnerabilidad.

Para almacenar toda la información obtenida mediante el escaneo, se han creado dos estructuras de datos en Python.

El primer tipo es el que se ha llamado Host que, como se puede ver a continuación (Código 1), tiene tres campos: el de la dirección IP, el del sistema operativo y el de las vulnerabilidades.

```
class Host(object):
    def __init__(self, ip=None, vulnerabilities=None, os=None):
        self.ip = ip
        self.vulnerabilities = vulnerabilities
        self.os = os
```

Código 1: Clase Host

En cuanto a la segunda clase o estructura de datos (Código 2), ésta tiene la función de almacenar toda la información relacionada con las vulnerabilidades y que es necesaria para hacer los cálculos del valor entre 1 y 10 del peligro de la vulnerabilidad. A continuación, se puede ver la clase creada para almacenar las vulnerabilidades.

```
class Vuln(object):
    def __init__(self, cve=None, score=None, vector=None, hasVector=None, cvss=None,
                 score_iot=None, vector_iot=None):
        self.cve = cve
        self.score = score
        self.vector = vector
        self.hasVector = hasVector
        self.cvss = cvss
        self.score_iot = score_iot
        self.vector_iot = vector_iot
```

Código 2: Clase Vuln

En este bloque del script sólo se llenarán los dos primeros campos de las vulnerabilidades, ya que los demás se llenarán en los próximos bloques. Primero encontramos el campo del CVE, y después tenemos el Score que es el valor entre 1 y 10 de la vulnerabilidad.

Una vez tenemos guardada toda la información necesaria, miramos vulnerabilidad por vulnerabilidad si hay alguna donde no haya guardado el CVE porque el escaneo no lo ha podido encontrar, y si se da el caso, borramos esa vulnerabilidad.

C. Consultas a la API de vulnerabilidades

Para obtener más información acerca de una vulnerabilidad, lo que se ha hecho ha sido utilizar una API que proporcionaba la organización NIST [13].

Con el fin de obtener información a través de la petición, sólo debe hacerse un simple GET. Por lo tanto, por cada petición se forma una URL, se hace un GET sobre ésta, y una vez nos hemos asegurado de que no ha devuelto ningún error mediante el *Status Code*, se transforma el resultado del GET en formato JSON. Un ejemplo de una petición es el siguiente:

https://services.nvd.nist.gov/rest/json/cve/1.0/CVE-2020-15778

Para medir el nivel de peligrosidad que tiene una vulnerabilidad, se utiliza un sistema de puntuación que mide diferentes aspectos de ésta llamados métricas. La entidad responsable de llevar a cabo este sistema de puntuación se llama **FIRST** (*Forum of Incident Response and Security Teams*) [14]. Y el sistema que tienen para puntuar las vulnerabilidades se llama **CVSS** (*Common Vulnerability Scoring System*) [15].

En este proyecto se calculará el rating final y se aplicarán los parámetros IoT sobre la versión CVSS 2.0. Desafortunadamente, la API que se utiliza para obtener el vector de cada vulnerabilidad siempre devuelve el vector de esta versión y, en muy pocos casos, devuelve el vector de la versión CVSS 3.0.

De todos modos, el código se ha adaptado para que, cambiando sólo un parámetro, pueda trabajar con la versión CVSS 3.0 (o su actualización 3.1). Porque si en un futuro esta API devuelve también los vectores de ambas versiones, el rating final se pueda obtener a partir de la versión CVSS 3.0 (o 3.1).

D. Aplicar parámetros IoT

Antes de aplicar los parámetros IoT, es necesario entender en qué consiste un vector CVSS.

El CVSS se compone de tres grupos de métricas diferentes; la Base, la Temporal y la Medioambiental (siendo las dos últimas opcionales para el cálculo mientras que la primera siempre es obligatoria).

Base Metric Group. Representa las características intrínsecas y fundamentales de una vulnerabilidad, que son constantes a lo largo del tiempo e independientes de los entornos y usuarios. En la Tabla 1 se muestra un resumen con las métricas del vector Base y sus correspondientes valores.

Tabla 1: Métricas CVSS 2.0

Base Metric Group		
Vector	**Tipus**	**Valor**
Access Vector (*AV*)	Local (*L*)	0.395
	Adjacent Network (*A*)	0.646
	Network (*N*)	1
Access Complexity (*AC*)	High (*H*)	0.35
	Medium (*M*)	0.61
	Low (*L*)	0.71
Authentication (*Au*)	Multiple (*M*)	0.45
	Single (*S*)	0.56
	None (*N*)	0.704
Confidentiality Impact (*C*)	None (*N*)	0
	Partial (*P*)	0.275
	Complete (*C*)	0.66
Integrity Impact (*I*)	None (*N*)	0
	Partial (*P*)	0.275
	Complete (*C*)	0.66
Availability Impact (*A*)	None (*N*)	0
	Partial (*P*)	0.275
	Complete (*C*)	0.66

Temporal Metric Group. Representa las características de una vulnerabilidad que cambia a lo largo del tiempo, pero que no depende de los usuarios.

Environmental Metric Group. Representa las características de una vulnerabilidad que cambia dependiendo de los usuarios.

Como el grupo temporal y medioambiental no son obligatorios para calcular el resultado final, no se han tenido en cuenta. Pero no sólo por este motivo, sino porque la API que se utiliza para obtener esta información, sólo devuelve los vectores del grupo Base.

A continuación, se muestran las fórmulas para calcular el rating de una vulnerabilidad. Estas fórmulas están sacadas directamente de la web de FIRST.

$$\begin{aligned}\text{BaseScore} = \text{roundToOneDecimal}(((0.6 \\ * \text{Impact}) + (0.4 \\ * \text{Exploitability}) - 1.5) \\ * \text{f (Impact)})\end{aligned} \quad (1)$$

$$\begin{aligned}\text{Impact} = 10.41 * (1 - (1 - \text{ConfImpact}) * (1 \\ - \text{IntegImpact}) * (1 \\ - \text{AvailImpact}))\end{aligned} \quad (2)$$

$$\begin{aligned}\text{Exploitability} = 20 * \text{AccessVector} \\ * \text{AccessComplexity} \\ * \text{Authentication}\end{aligned} \quad (3)$$

$$\begin{aligned}\text{f (impact)} = 0 \text{ si Impact} = 0, \\ \text{sino f (impact)} = 1.176\end{aligned} \quad (4)$$

Para hacer los cálculos con el CVSS 3.1, nos basamos en la Tabla 2, la cual muestra diferencias con la tabla de la versión CVSS 2.0 (Tabla 1).

Tabla 2: Métricas CVSS 3.1

Base Metric Group		
Vector	**Tipus**	**Valor**
Access Vector (*AV*)	Local (*L*)	0.55
	Adjacent Network (*A*)	0.62
	Network (*N*)	0.85
	Physical (*P*)	0.2
Access Complexity (*AC*)	High (*H*)	0.44
	Low (*L*)	0.77
Privileges Required (*PR*)	High (*H*)	0.27 (*0.5 si Scope = C*)
	Low (*L*)	0.62 (0.68 si Scope = *C*)
	None (*N*)	0.85
User Interaction (*UI*)	Required (*R*)	0.62
	None (*N*)	0.85
Scope (*S*)	Unchanged (*U*)	N/A
	Changed (*C*)	N/A
Confidentiality Impact (*C*)	None (*N*)	0
	Low (*L*)	0.22
	High (*H*)	0.56
Integrity Impact (*I*)	None (*N*)	0
	Low (*L*)	0.22
	High (*H*)	0.56
Availability Impact (*A*)	None (*N*)	0
	Low (*L*)	0.22
	High (*H*)	0.56

Las fórmulas que se usan para calcular el rating de una vulnerabilidad en la versión CVSS 3.0 son las siguientes:

$$\text{Si Impact} \leq 0\text{, BaseScore} = 0 \quad (5)$$

Si *Scope* no cambia (Unchanged):

$$\text{BaseScore} = \text{RoundToOneDecimal}(\text{Minim}[(\text{Impact} + \text{Exploitability}), 10] \quad (6)$$

Si cambia:

$$\text{BaseScore} = \text{RoundToOneDecimal}(\text{Minim}[1.08 * (\text{Impact} + \text{Exploitability}), 10] \quad (7)$$

Para calcular la *Exploitability*:

$$\text{Exploitability} = 8.22 * \text{AttackVector} * \text{AttackComplexity} * \text{PrivilegesRequired} * \text{UserInteraction} \quad (8)$$

Finalmente, para el *Impact*, si *Scope* no cambia (Unchanged):

$$\text{Impact} = 6.42 * \text{ISS} \quad (9)$$

Si cambia:

$$\text{Impact} = 7.52 * (\text{ISS} - 0.029) - 3.25 * (\text{ISS} - 0.02) \quad (10)$$

dónde

$$\text{ISS} = 1 - [(1 - \text{Confidentiality}) * (1 - \text{Integrity}) * (1 - \text{Availability})]$$

Para adaptar estos cálculos a un entorno IoT, se utiliza la información de [16] dónde se explica qué cambios son necesarios para que el resultado esté adaptado a un entorno IoT. Se puede integrar muy fácilmente al proyecto ya que sólo se modifican pesos de las métricas y se añade un parámetro llamado *Human Safety Index*. A continuación, se explican con detalle estos cambios.

Cambio en los valores del vector AV. Se propone añadir dos valores nuevos enfocados a un entorno IoT, es decir, se mantendrá el valor L y se añadirá el valor L_i con una métrica de 0.6. Y para el valor P, se añadirá el valor P_i con una métrica de 0.44. Estas métricas son ligeramente más altas ya que es más fácil acceder físicamente a los dispositivos IoT.

Cambio en los valores del vector AC. En este caso se añade un valor nuevo M con una métrica de 0.44 y de la misma manera que en el caso anterior, aparte del valor H, se añade un valor H_i con una métrica de 0.2. Como se requiere más complejidad para llevar a cabo un ataque a los dispositivos IoT, se les asigna unas métricas ligeramente inferiores en comparación a las métricas definidas para dispositivos convencionales.

Human Safety Index. Se añade al grupo Base y Environmental y mide el nivel de seguridad para los humanos, ya que si, por ejemplo, tenemos maquinaria conectada a Internet en una fábrica, un ataque a estas máquinas podría provocar daños humanos que se deben tener en cuenta de cara al resultado final.

Las fórmulas siguen siendo las mismas que antes, pero con los nuevos valores de las métricas y añadiendo el *Human Safety Index* al cálculo del *Impact*. Es importante remarcar que el valor del *Human Safety Index* en caso de ser 0 no afecta a los cálculos, ya que como se puede ver en la siguiente fórmula, el valor sólo está para darle más precisión a los cálculos en caso de ser diferente a 0.

$$\text{ISS} = 1 - [(1 - \text{Confidentiality}) * (1 - \text{Integrity}) * (1 - \text{Availability}) * (1 - \text{HSI})] \quad (11)$$

Para que todo el proceso se entienda mejor, se muestra el cálculo para un entorno convencional y un entorno IoT teniendo en cuenta una misma vulnerabilidad.

Se utiliza como ejemplo la vulnerabilidad CVE-2020-15778. Empezamos listando el vector en un entorno convencional y en un entorno IoT.

- Vector: AV:N/AC:M/Au:N/C:P/I:P/A:P
- Vector IoT: AV:N/AC:Mi/Au:N/C:P/I:P/A:P/Hi:Ni

El cálculo en un entorno convencional sería:

$$\text{Impact} = 10.41 * \left(1 - (1 - 0.275) * (1 - 0.275) * (1 - 0.275)\right) = 6.44 \quad (12)$$

$$\text{Exploitability} = 20 * 1 * 0.61 * 0.704 = 8.58 \quad (13)$$

$$f(\text{impact}) = 0 \text{ si Impact} = 0, \text{ sino } f(\text{impact}) = 1.176 \quad (14)$$

$$\text{BaseScore} = \text{roundToOneDecimal}\left(((0.6 * 6.44) + (0.4 * 8.58) - 1.5) * 1.176\right) = \mathbf{6.8} \quad (15)$$

En cambio, para un entorno IoT tendríamos:

$$\text{Impact} = 10.41 * \left(1 - (1 - 0.275) * (1 - 0.275) * (1 - 0.275) * (1 - 0)\right) = 6.44 \quad (16)$$

$$\text{Exploitability} = 20 * 0.85 * 0.44 * 0.704 = 5.27 \quad (17)$$

$$f(\text{impact}) = 0 \text{ si Impact} = 0, \text{ sino } f(\text{impact}) = 1.176 \quad (18)$$

$$\text{BaseScore} = \text{roundToOneDecimal}\Big(((0.6 * 6.44) + (0.4 * 5.27) - 1.5) * 1.176\Big) = \mathbf{5.2} \quad (19)$$

E. Cálculo del rating final

Para calcular el rating final de la red, se hace la media aritmética de los *scores* de todas las vulnerabilidades de todos los dispositivos.

Además, el script hace la media tanto en un entorno IoT como en uno convencional para poder comparar los resultados. La fórmula utilizada en este caso es la siguiente:

$$\text{Score} = \text{RoundToOneDecimal}\Big(\sum_{i=0}^{numVuln} \frac{numVuln_i}{i}\Big) \quad (20)$$

Este método no es del todo preciso ya que no todos los dispositivos de la red tienen la misma importancia. Por ejemplo, no es lo mismo una vulnerabilidad en el router que está entre Internet y la red interna, que una vulnerabilidad en una bombilla inteligente.

En este proyecto se ha hecho una aproximación de este riesgo final, ya que no se ha implementado ningún sistema de pesos que pueda dar más importancia a ciertas vulnerabilidades dependiendo de en qué dispositivo se encuentren.

F. Resultados

El script se ha lanzado en una red de laboratorio disponible en la propia universidad con diferentes dispositivos los cuales tienen varias vulnerabilidades. Los resultados se resumen en la Tabla 3 y Tabla 4.

Tabla 3: Resultados obtenidos por host

Host	Score	Score IoT	Num. Vulnerabilidades
1	4.9	4.8	41
2	5.3	5.1	33
3	5.1	5	35
4	4.5	4.4	23
5	5.5	5.1	51
6	5.1	4.4	28
7	5.6	5.1	20
8	6.9	6.2	14
9	0	0	0
10	0	0	0
11	5.6	4.9	95
12	5.2	4.8	58
13	5	4.7	43
14	0	0	0

Tabla 4: Resultados totales

	Score	Score IoT	Num. Vulnerabilidades
Total	5.3	4.9	441

Para tener una representación más visual, se han generado unas gráficas del resultado obtenido. En la Fig. 3 se muestra la comparativa del valor medio del *score* tanto en un entorno convencional como en uno IoT para las diferentes vulnerabilidades de cada host. Se puede apreciar que, exceptuando el host 9, 10 y 14, el script ha detectado vulnerabilidades. En color azul el score medio que tendrían las vulnerabilidades si el host estuviera en un entorno convencional, y en color verde si estuviera en un entorno IoT.

Fig. 3. Comparativa de las vulnerabilidades en ambos entornos

La siguiente gráfica, Fig. 4, muestra la media del score de todas las vulnerabilidades de todos los hosts, es decir, a partir del valor medio de cada host se ha obtenido un score final, el cual en color azul sería de un entorno convencional y en color verde de un entorno IoT.

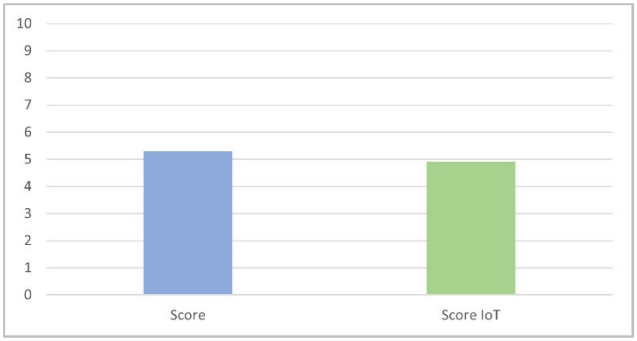


Fig. 4. Score final de la red

Además de estas gráficas, el propio script genera un *report* con información más detallada de cada vulnerabilidad y de cada dispositivo.

IV. CONCLUSIONES

Después de haber desarrollado la automatización de un análisis de riesgos en un entorno IoT, se ha visto la utilidad de los métodos que se utilizan para calcular la severidad de

las vulnerabilidades con relación a otras, ya que gracias a estos cálculos uno se puede hacer una idea del riesgo actual de su red o dispositivo. En este proyecto se acabó utilizando la metodología CVSS ya que es la más conocida y estandarizada actualmente.

Los resultados han sorprendido un poco, ya que el resultado obtenido de la misma vulnerabilidad en un entorno IoT, generalmente ha sido más bajo que el resultado obtenido en una red convencional. Esto se debe a que después de aplicar los parámetros específicos para entornos IoT, se obtenía un vector mucho más preciso en cuanto a la severidad de la vulnerabilidad, haciendo que, con este aumento de precisión, el resultado variará respecto al valor inicial.

Estos parámetros son el *Attack Vector (AV), el Attack Complexity (AC) y el Human Safety Index*. Se han modificado específicamente para satisfacer requisitos del IoT que en el CVSS original no se tienen en cuenta.

Al analizar los resultados obtenidos vemos que tienen su lógica. Esto se debe a cómo se han aplicado los pesos de cada vector, es decir, el hecho de haberlos aplicado de manera estática, hace que se traten todas las vulnerabilidades por igual, por lo tanto, no se diferencia entre si una vulnerabilidad puede afectar más por el simple hecho de ser un entorno IoT o no.

Todas las vulnerabilidades se comportan diferente, y sobre todo si el entorno cambia, de modo que aplicar valores estáticos no es del todo preciso. Según la información extraída de [16], al valor de *Human Safety* como no tiene ninguna referencia previa en el vector original y es un campo que se añade nuevo, se le asigna el valor de 0 para todas las vulnerabilidades (para que no afecte al cálculo final). En cambio, al variar el valor del AV y AC, vemos como el resultado final es distinto, haciendo que sea más bajo.

V. Líneas de futuro

A continuación, se listarán las líneas de futuro o mejoras que se pueden aplicar a esta primera aproximación del bloque descrito.

- Usar CVSS 3.0 en vez de CVSS 2.0.
- Asignar los pesos de forma dinámica. Realizar la asignación de más o menos peso dependiendo de la vulnerabilidad.
- Asignar el *Human Safety Index* de manera dinámica, asignando un valor distinto dependiendo del tipo de dispositivo y su criticidad. Para ello se podría generar una tabla de valores dónde se clasificarían los dispositivos IoT por tipo, funcionalidad o criticidad.
- Calcular la media final teniendo en cuenta la criticidad de un dispositivo en una red, por ejemplo, dar más importancia a las vulnerabilidades de un router que las vulnerabilidades de un sensor cualquiera.
- Para terminar, se podría generar un informe con más detalles de cada vulnerabilidad y dispositivo.

Por otro lado, recordar que los próximos pasos deben contemplar la implementación de los otros bloques que componen el sistema descrito en la Fig. 1.

Referencias

[1] Sánchez, J., Corral, G., de Pozuelo, R. M., & Zaballos, A. (2016). Security issues and threats that may affect the hybrid cloud of FINESCE. Netw. Protoc. Algorithms, 8(1), 26-57.

[2] SPRINT 4.0 Project, 2020, [Online] “https://www.sprint40.eu/

[3] Zaballos, A., Briones, A., Massa, A., Centelles, P., & Caballero, V. (2020). A smart campus’ digital twin for sustainable comfort monitoring. Sustainability, 12(21), 9196.

[4] PublicaTIC, “Controles y auditoría del IoO,” 2019, [Online] https://blogs.deusto.es/master-informatica/controles-y-auditoria-del-iot/

[5] INCIBE, “La importancia de la seguridad en iot. principales amenazas,” 2019, [Online] https://www.incibe-cert.es/blog/importancia-seguridad-iot-principales-amenazas

[6] OWASP, “Top 10 web application security risks,” 2019, [Online] https://owasp.org/www-project-top-ten/

[7] ETSI, “ETSI EN 303 645: Cyber Security for Cunsomer Internet of Things: Baseline Requierements,” 2020, [Online] https://www.etsi.org/deliver/etsi_en/303600_303699/303645/02.01.01_60/en_303645v020101p.pdf

[8] C. Otero, “Cómo configurar con más seguridad tus dispositivos IoT y red casera,” 2019, [Online] https://as.com/meristation/2019/09/06/betech/1567723080_550087.html

[9] NIST, “NISTIR 8259: Foundational Cybersecurity Activities for IoT Device Manufacturers” 2021, [Online] https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8259.pdf

[10] G. Lyon, “Nmap.org,” 2020, [Online] https://insecure.org/fyodor/

[11] TOKYONEON, “Easily detect cves with nmap scripts,” 2019, [Online] https://null-byte.wonderhowto.com/how-to/ easily-detect-cves-with-nmap-scripts-0181925/

[12] Tenable, “Nessus es la solución n.° 1 para evaluaciones de vulnerabilidades,” 2020, [Online] https://es-la.tenable.com/products/Nessus

[13] B. Byers and H. Owen, “Automation support for cve retrieval,” 2019, [Online] https://csrc.nist.gov/CSRC/media/Projects/National-Vulnerability-Database/documents/web%20service%20documentation/Automation%20Support%20for%20CVE%20Retrieval.pdf

[14] FIRST, “First is the global forum of incident response and security teams,” 2020, [Online] https://www.first.org/

[15] FIRST., “Common vulnerability scoring system version 3.1: Specification document,” 2019, [Online] https://www.first.org/cvss/specification-document

[16] A. Ur-Rehman, I. Gondal, J. Kamruzzaman, and A. Jolfaei, “Vulnerability modelling for hybrid it systems.” in ICIT, 2019, pp. 1186–1191.


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

eHDDP: enhanced Hybrid Domain Discovery Protocol for network topologies with both wired/wireless and SDN/non-SDN devices

Isaias Martinez-Yelmo, Joaquin Alvarez-Horcajo, Juan A. Carral, Diego Lopez-Pajares

Universidad de Alcalá, Departamento de Automática, 28805 Alcalá de Henares, Spain.
{isaias.martinezy, j.alvarez, juanantonio.carral, diego.lopezp}@uah.es

Resumen

Handling efficiently both wired and/or wireless devices in SDN networks is still an open issue. eHDDP allows the SDN control plane to discover and manage hybrid topologies composed of both SDN and non-SDN devices with wired and/or wireless interfaces, thus opening a path for the integration of IoT and SDN networks. eHDDP has been thoroughly evaluated in different scenarios and exhibits good scalability properties since the number of required messages is proportional to the number of existing links in the network topology. Moreover, the obtained discovery and processing times are small enough to support scenarios with low mobility devices since they are in the range of hundreds of milliseconds.

***Key Words*—Hybrid, SDN, WSN, IoT, Discovery, unidirectional, bidirectional**

I. SUMMARY

This submission presents enhanced Hybrid Domain Discovery Protocol (eHDDP), a work previously published in [1], it is an enhanced version of Hybrid Domain Discovery Protocol (HDDP) [2], which allows its use not only in wired interfaces but also in wireless interfaces, including low powered wireless technologies, in an efficient way. eHDDP is the first proposal offering a general solution to the problem of discovering fully hybrid SDN network topologies comprised of both Software Defined Networking (SDN) and non-SDN devices with wired and/or wireless interfaces to the best of our knowledge. Some interesting use cases of the discovery of wireless hybrid SDN networks are the integration of Wireless Sensor Networks (WSNs) or the IoT networks by the SDN control plane to manage and supervise them in a more accurate way. The main contributions of this work published in [1] are:

- The upgrade of the HDDP header in the control messages to support different types of wireless interfaces. The header fields have been revised to minimize its size to support technologies with a small Maximum Transfer Unit (MTU) size.
- The definition of a protocol logic for wireless interfaces. This new logic considers the number of neighbours is unknown in a wireless interface since it may change with time because of mobile devices.
- The detection of both unidirectional and bidirectional links in wireless scenarios.
- A reference implementation of eHDDP comprised of a network agent in non-SDN devices and an application in the SDN control plane by using Open Network Operating System (ONOS) and the OpenFlow protocol.
- The enhancement of previously existing network emulation tools to properly deploy and perform experiments with wireless Ad-Hoc networks with multiple devices.
- A thorough evaluation of eHDDP under different types of network topologies: wired, wireless and both wired and wireless topologies.

ACKNOWLEDGEMENT

This work was funded by grants from Comunidad de Madrid through Project TAPIR-CM (S2018/TCS-4496) and Project IRIS-CM (CM/JIN/2019-039), from Junta de Comunidades de Castilla la Mancha through Project IRIS-JCCM (SBPLY/19/180501/000324).

REFERENCES

[1] Joaquin Alvarez-Horcajo, Elisa Rojas, Isaias Martinez-Yelmo, Marco Savi, Diego Lopez-Pajares, "HDDP: Hybrid Domain Discovery Protocol for Heterogeneous Devices in SDN", Computer Networks, vol. 191, id 107983, May. 2021

[2] Isaias Martinez-Yelmo, Joaquin Alvarez-Horcajo, Juan A. Carral, Diego Lopez-Pajares, ."HDDP: Enhanced Hybrid Domain Discovery Protocol for network topologies with both wired/wireless and SDN/non-SDN devices", IEEE Communications Letters, vol. 24, no. 8, 1655-1659, Aug. 2020


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Planificaciones de enlace basadas en estimación del canal aplicadas a 5G

Jose J. Rico-Palomo*, Jesús Galeano-Brajones*, Juan F. Valenzuela-Valdes†, David Cortes-Polo*, Javier Carmona-Murillo*

* Departamento de Ingeniería de Sistemas Informáticos y Telemáticos. Universidad de Extremadura. 06006.
† Departamento de Teoría de la señal, Telemática y Comunicaciones. Universidad de Granada. 18014.

jjricopal@unex.es, jgaleanobra@unex.es, juanvalenzuela@ugr.es, dcorpol@unex.es, jcarmur@unex.es

El crecimiento exponencial de dispositivos conectados a Internet y el aumento de los volúmenes de tráfico que se generan dificultan una gestión eficiente de las conexiones entre las estaciones base y los usuarios en las redes móviles. Esto desemboca en una sobrecarga y en un mayor consumo de tiempo y recursos en la red de acceso. Para el estudio de estos problemas, se clasifican los servicios de los usuarios según los requerimientos del tipo de tráfico (capacidad, latencia, fiabilidad, densidad de conexiones, etcétera). En este trabajo se proponen varios mecanismos que, mediante el envío y la recepción de pilotos por parte del terminal móvil, permiten estimar el enlace en busca de conexiones con mayor capacidad y/o menor latencia. Estos mecanismos descargan a la red de la toma de decisiones, mejorando la planificación de recursos según el tipo de tráfico que genere el usuario. Los resultados obtenidos demuestran una disminución y estabilización de la latencia, así como un aumento en la capacidad media de los enlaces. Se reducen, además, las conexiones bloqueadas por falta de recursos en la red.

***Palabras Clave*—5G, redes celulares, estimación de canal, capacidad, latencia, URLLC, eMBB, mMTC.**

I. INTRODUCCIÓN

Desde comienzos de 2020 el número de teléfonos móviles inteligentes ha aumentado en 93 millones, siendo actualmente la cifra de dispositivos conectados superior a 5,22 miles de millones en todo el mundo. Este aumento de dispositivos se traduce inevitablemente en un aumento del tráfico de datos, que ya el año 2021 ha registrado un volumen superior a 55 Exabytes al mes [1]. La evolución en los últimos años de las aplicaciones más utilizadas, como la Realidad Virtual/Aumentada (AR *Augmented reality*, VR *Virtual Reality*), los vehículos autónomos y conectados, los videojuegos en línea, el Internet de las Cosas (IoT, *Internet of Things*) y las plataformas de vídeo bajo demanda, entre otras, han contribuido a este aumento exponencial. Si a esto le sumamos el crecimiento esperado de estas aplicaciones, que según [2] será de un 175% en 10 años, se muestra un escenario que sobrepasa las especificaciones que ofrecen las redes convencionales.

5G, la nueva generación de redes celulares, nace bajo la motivación de dar soporte a todas las conexiones existentes, para cubrir la demanda de altas tasas de datos y los requerimientos de distintos tipos de comunicación, como latencia, número de usuarios conectados, fiabilidad, etcétera. Estas redes utilizan una combinación de tecnologías habilitadoras de nivel físico, como las antenas MIMO (*Multiple-Input Multiple-Output*) masivo, codificación digital para el *beamforming* y despliegues ultra densos de celdas, entre otros, junto con tecnologías de nivel de red, como SDN (*Software-Defined Networking*), NFV (*Network Function Virtualization*), MEC (*Mobile Edge Computing*), etcétera. Una de las aplicaciones más directas de la combinación de estas tecnologías es asignar recursos a los terminales conectados de forma eficiente dependiendo del tipo de tráfico, de manera que no se sobredimensionen los enlaces pero que puedan tener recursos suficientes para dar servicio a la demanda. Este proceso de asignación se realiza, normalmente, antes del establecimiento de la conexión, cuando el usuario y la red acuerdan los requerimientos de la comunicación, pero una vez evaluado el nivel de señal-ruido-interferencia (SINR, *Signal-to-Interference-plus-Noise-Ratio*) del usuario a cada una de las estaciones base candidatas. Esto consume unas unidades de tiempo que, dependiendo de la criticidad de la aplicación (por ejemplo, vehículos autónomos), pueden ser determinantes, ya que aumenta el tiempo de respuesta de la red a las peticiones. Además, se genera una alta sobrecarga en la red de acceso (RAN, *Radio Access Network*) en la toma de decisión para cada una de

las conexiones entrantes.

Con estos nuevos paradigmas y con el aumento de la complejidad de las aplicaciones y dispositivos móviles, surge una nueva manera de clasificar los tipos de servicio, no basándose únicamente en sus casos de uso, si no en los requerimientos que necesitan para satisfacer sus demandas. Esta clasificación divide a los servicios en tres grandes bloques [3]:

- *Enhanced Mobile Broadband* (eMBB): requerimientos de altas tasas de datos, gran ancho de banda y grandes zonas de cobertura. En este bloque se englobarían, por ejemplo, las aplicaciones de vídeo bajo demanda y de AR/VR.
- *Ultra-Reliable and Low Latency Communications* (URLLC): comunicaciones de muy baja latencia extremo a extremo y una fiabilidad superior al 99'9%. Las comunicaciones críticas, como los vehículos autónomos, entran dentro de esta clasificación.
- *Massive Machine-Type Comunications* (mMTC): estas aplicaciones (como, por ejemplo, las de industria 4.0 e IoT) tienen bajas tasas de datos pero en numerosos dispositivos, por lo que requieren una alta densidad de conexiones, restándole importancia a otros KPIs (*Key Performance Indicators*) como la latencia.

Ya en la literatura se proponen mecanismos que ayudan a gestionar de manera eficiente cada tipo de conexión. En [4] se plantean mecanismos basados en planificación de enlace para comunicaciones eMBB y URLLC donde se establece la fiabilidad y la pérdida de paquetes como KPIs a evaluar. Algunos trabajos presentan soluciones en el RAN basadas en optimización e inteligencia artificial [5], [6] para reservar los recursos necesarios de capacidad y/o latencia que necesiten los usuarios según sus requerimientos. Otras soluciones basadas en *RAN-slicing* plantean segmentar la red de acceso en divisiones lógicas, cada una de ellas preparadas para dar un servicio eficiente a cada tipo de usuario, compartiendo los recursos disponibles en toda la red. Estas soluciones están basadas en software, haciendo recaer el cómputo en las estaciones base [7], [8].

Siguiendo la tendencia de los trabajos comentados previamente, en este artículo se presenta una solución basada en la estimación del canal mediante el uso de pilotos, donde se libera a la RAN de la toma de decisiones y se delega esta tarea a los terminales móviles. Estos mecanismos de estimación prescinden del establecimiento de conexión para escoger la estación base que mejores métricas ofrezca, centrándose directamente en la evaluación de latencia o capacidad que pueda servir la estación base.

El resto del trabajo está organizado con la siguiente estructura: en la Sección II se presenta la formalización matemática utilizada para modelar el sistema y su comportamiento; la Sección III describe los mecanismos implementados para la estimación del canal; la Sección IV presenta las pruebas de simulación realizadas y los resultados obtenidos; por último, en la Sección V, se presentan las conclusiones.

II. MODELADO DEL SISTEMA

A continuación se detallan los elementos más importantes que influyen en los cálculos a lo largo del tiempo de simulación y los KPIs que interesan para la obtención de resultados y su modelización formal.

A. Planificación de enlace basada en SINR

Cuando un terminal móvil necesita conectarse al RAN, se utiliza la SINR para evaluar cada una de las estaciones base candidatas. En la generalidad de los casos, la estación base a la que se conectará será aquella que mejor nivel de SINR tenga (medido en dBm). Los cálculos se realizan siguiendo la siguiente ecuación [9]:

$$SINR_u = \frac{P_{RX(i,u)}(mW)}{[\sum_{j=1 j\neq i}^{N} P_{RX(j,u)}(mW)] + P_{N_0}(mW)} \quad (1)$$

donde $P_{RX(i,u)}$ es la potencia recibida por el usuario u de cada una de las estaciones base, en miliwattios; P_{N_0} es la potencia de ruido y $[\sum_{j=1 j\neq i}^{N} P_{RX(j,u)}(mW)]$ es la interferencia, es decir, la suma de la potencia recibida de cada estación base distinta a la evaluada, pero que operan en la misma frecuencia.

B. Canal de propagación y modelo de pérdidas

El enlace entre un transmisor y un receptor se evalúa mediante la formula del presupuesto de potencia o *link budget*, que se usa también para calcular la potencia recibida en un nodo móvil teniendo en cuenta las propiedades de las antenas y las pérdidas en espacio libre. La fórmula se presenta a continuación:

$$\begin{aligned} P_{RX}(dBm) = P_{TX}(dBm) + G_{TX}(dBm) + \\ + G_{RX}(dBm) - L(dB) \end{aligned} \quad (2)$$

La potencia recibida, denotada por P_{RX} es combinación entre la potencia transmitida, P_{TX}, las ganancias de las antenas transmisoras y receptoras, G_{TX} y G_{RX} respectivamente, y las pérdidas en espacio libre L.

Para el cálculo de las pérdidas en espacio libre se ha utilizado el modelo ABG, utilizado en numerosos trabajos técnicos y académicos [10]. Las pérdidas en espacio libre se calculan en función de la frecuencia de radiación y la distancia entre el transmisor y receptor, entre otros factores:

$$\begin{aligned} PL^{ABG}(f,d)[dB] = 10\alpha \log_{10}\left(\frac{d}{1\text{m}}\right) + \beta + \\ + 10\gamma \log_{10}\left(\frac{f}{1\text{GHz}}\right) + X_\sigma^{ABG} \end{aligned} \quad (3)$$

donde $PL^{ABG}(f,d)$ son las pérdidas en espacio libre en decibelios para la distancia d y una frecuencia determinada

f. α y β y γ son los coeficientes propios del modelo, que varían en función del canal de propagación en el que se encuentre el enlace. Estos valores son medidos en circunstancias concretas para estimar las pérdidas [11].

C. Modelo de capacidad de antenas MIMO

Las antenas tanto de los terminales móviles como de los paneles de las estaciones base se modelan con una tecnología de antenas MIMO con diversidad de antenas tanto en la recepción como en la transmisión. La capacidad de un enlace de esta tecnología es el producto entre el ancho de banda asignado para el enlace por la eficiencia espectral entre las dos antenas, medida en Mbps. La eficiencia espectral (medida en bits/s/Hz) depende del canal de propagación y del número de antenas transmisoras y receptoras, siguiendo la fórmula que se muestra a continuación [12]:

$$S_c = \log_2\left(det\left[I_{N_{RX}} + \frac{SINR}{N_{TX}} H * H^{T'}\right]\right) \quad (4)$$

donde $I_{N_{RX}}$ es la matriz de identidad cuyas dimensiones son el número de antenas MIMO del receptor, N_{TX} es el número de antenas del transmisor y H es la matriz del canal. $H^{T'}$ es la transposición conjugada de esta matriz.

D. Modelo de latencia

El modelo de latencia utilizado [13] es composición de 3 tiempos que afectan a la comunicación entre el terminal móvil y la estación base en el enlace *uplink*.

$$T_{total} = \text{link latency} = T_{prop} + T_{tail} + T_{hand} \quad (5)$$

Los tiempos de cola $T_{tail} = \frac{\beta}{\mu(1-\beta)}$ y procesamiento $T_{hand} = \frac{1}{\mu(1-\beta)}$ son los correspondientes al tiempo que se consume al estar encolada la conexión en la estación base y al tiempo que tarda el *NodeB* en procesar la demanda, respectivamente. Ambos atienden a un modelo de cola $GI|M|1$ donde los tiempos de servicio se distribuyen exponencialmente con ratio μ. El tiempo de propagación T_{prop} corresponde al tiempo que tarda la propagación de la información por el medio, desde el momento en el que se emite hasta que llega al receptor. Este tiempo está fuertemente condicionado por las propiedades del canal. La fórmula utilizada para modelar esta latencia es la siguiente:

$$T_{prop} = \frac{2(t_{slot} - E[T_v])}{1 + f_{err}\left(\frac{\delta(d,f)}{\sqrt{2\sigma}}\right)} \quad (6)$$

donde t_{slot} es el tiempo entre los PRBs (*Physical Resource Block*), definido por el estándar 5G [14], f_{err} es la función error y $\delta(f,d) = P_{TX} + P_{N_0} - L(f,d)$. La dependencia con el canal de propagación se refleja en $E[T_v]$, que define las alteraciones que producen los bloqueos móviles. Los bloqueos móviles son todo tipo de objetos que interfieren en la línea de visión del propio enlace, como peatones, vehículos, etcétera. Estos bloqueos se modelan siguiendo una modelo de cola $M|G1|\infty$, donde los instantes de llegada son interpretados como el cruce del campo de visión durante un instante de tiempo, donde cambiará de LOS (*Line-Of-Sight)* a NLOS (*No-Line-Of-Sight)*. Con los tiempos medios en el que enlace está visible y bloqueado se calcula la alteración del enlace [15].

E. Envío y recepción de pilotos

La forma escogida de estimar el canal es introducir un símbolo piloto conocido en la señal de transmisión, mediante la propia modulación de la señal. Los pilotos son señales usadas para la supervisión, control, ecualización y/o sincronización del canal, entre otros propósitos. Esa señal conocida por emisor y receptor viajará por el canal sufriendo una alteración. Cuando el receptor evalúe el piloto lo comparará con el valor conocido y podrá estimar la degradación sufrida, para conocer así el estado del canal y alguna de sus propiedades [16].

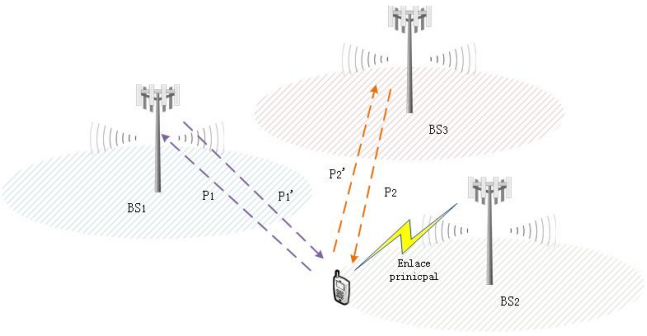


Fig. 1. Proceso de envío y recepción de pilotos. El terminal móvil, sin descartar el enlace principal ni parar el flujo de información, envia y recibe continuamente pilotos de las demás estaciones base cercanas.

En el sistema simulado, los pilotos son enviados por el nodo móvil a la estación base, y son devueltos por la misma. Esto ayuda a mejorar la precisión de la estimación, realizando dos estimaciones de canal (*downlink* y *uplink*) en vez de sólo una como se describió anteriormente (sólo *uplink*).

El nodo móvil envía estos pilotos a todas las estaciones base en su rango de cobertura, y evalúa los recibidos. La selección de la estación base no se hace, por tanto, por nivel de SINR, sino por la propiedad del canal que más interese en cada momento.

F. Modelos de tráfico y demanda

Para distinguir los distintos tipos de servicios y aplicaciones, se implementan varios modelos de tráfico que aproximan el tamaño, la duración y la tasa de llegadas del tráfico de varias aplicaciones reales. Este modelado matemático del teletráfico se basa fundamentalmente en el trabajo realizado en [17]. Los modelos implementados se resumen en la Tabla I.

Los nodos móviles pre-simulan el tráfico que generarán a lo largo del tiempo de ejecución, y crean las demandas. Estas estructuras corresponden a la cantidad de recursos (capacidad en Mbps) que serán ocupados a lo largo de la red: las estaciones base por las que se enrutará la conexión hasta llegar a la red de acceso y sus enlaces.

Aplicación	Referencia	Tipo	Modelo
FTP	[18]	Estocástico	Truncated lognormal distribution
Gaming	[19], [20]	Experimental	Volumen de tráfico y tasas de llegada reales
VOD	[21], [22]	Mixto	Truncated pareto distribution para las de llegada con volúmenes de tráfico reales
Web	[21]	Estocástico	Truncated lognormal distribution
IoT	[23]	Estocástico	Poisson process
AR/VR	[24], [25]	Experimental	Volumen de tráfico y tasas de llegada reales
V2X	[26], [27]	Mixto	Mezcla entre tráfico de control, datos y VOD, entre otros

Tabla I
MODELOS DE TRÁFICO IMPLEMENTADOS EN EL FRAMEWORK DE SIMULACIÓN

III. MECANISMOS PROPUESTOS

En esta sección se presentan los mecanismos propuestos para la planificación de enlace de terminales móviles según las necesidades del tráfico que generan y la criticidad de su comunicación. Si un usuario necesita menos latencia o más capacidad de las que pueden permitir sus enlaces con la red, necesitará utilizar otro tipo de planificación de enlace que le permita conectarse a otra estación base que sí pueda proporcionarle dichos recursos. Estas soluciones están basadas en la estimación del canal, que proporcionan al terminal móvil una visión completa del estado de la red según las dos métricas a evaluar, delegando el cómputo de la estimación en el propio terminal y no en la red. Se proponen a continuación dos mecanismos.

A. Planificación de enlace basada en estimación de latencia

Para estimar la latencia del enlace, el terminal móvil debe enviar una señal piloto conocida a las estaciones base, que retornarán la señal. El tiempo que tarda el piloto entre el envío y su posterior recepción puede ayudar a estimar la latencia de propagación t_{prop}; pero el usuario sólo conoce el tiempo total (bidireccional). Para calcular el tiempo sólo en una dirección, es necesario desglosar la escena:

$$t_{pilot} = t_{prop}(P_{TX_{UE}}, f_{up}) + t_{hand} + t_{BS}(P_{TX_{BS}}, f_{down}) \quad (7)$$

Este tiempo total t_{pilot} es la combinación entre varios tiempos: el tiempo que tarda la señal piloto en llegar a la estación base t_{prop}. Según el modelado del sistema, y teniendo en cuenta que el canal es común entre los enlaces *uplink* y *downlink* (ya que se entiende que el tiempo que se consume para la interacción del piloto es despreciable en comparación con el tiempo de coherencia del canal), este tiempo depende de la potencia de transmisión del terminal $P_{TX_{UE}}$ y la frecuencia del enlace ascendente f_{up}. El tiempo que tarda la señal piloto en ser enviada por la estación base y llegar nuevamente al terminal t_{BS} sigue el mismo esquema, siendo dependiente de la potencia de transmisión de la estación base $P_{TX_{BS}}$ y la frecuencia del enlace ascendente f_{down}.

Si consideramos despreciable el tiempo que tarda la estación base en procesar la señal t_{hand} y conocidas las marcas de tiempo de envío y recepción del piloto por parte del terminal móvil, además del tiempo de retorno de las mismas pero por parte de la estación base, la latencia del enlace puede ser estimada mediante la siguiente fórmula:

$$L'_{j,i} = t_{pilot}(UE_j, BS_i) - T'_{prop}(BS_i, UE_j, f_{down}), \quad \forall i \in \{BN_{BS}\}, \forall j \in \{U\} \quad (8)$$

B. Planificación de enlace basada en estimación de capacidad

El modelado de la capacidad, como se ha desglosado anteriormente, es función del número de transmisores N_{tx} y receptores N_{rx}, el ancho de banda BW_u y las propiedades del canal H. Para la estimación de la capacidad del terminal $C'(BW_u, SINR, N_{tx}, N_{rx}, H)$, el usuario debe conocer algunas características del sistema receptor, como el ancho de banda y el número de antenas MIMO del extremo del enlace.

Las señales pilotos proporcionarán esa información mediante el mismo mecanismo de envío y recepción, devolviendo cierta información de la estación base: el número de receptores MIMO que habilitará para el enlace y el ancho de banda que se le asignaría al usuario en el caso de que se conectase, sin tener en cuenta ningún algoritmo de asignación inteligente, puesto que a priori la estación base no conoce las características del tráfico del usuario ni la criticidad de su comunicación.

Para estimar las condiciones del canal, por tanto, es necesario generar aleatoriamente matrices H utilizando una distribución normal compleja $\mathcal{N}(\mu, \sigma^2)$, según los parámetros que se detecten en la recepción de la señal piloto:

$$\mathbf{H'} = \sum^{N_H} \sum^{M} \frac{\mathbf{H}_{1(N_{tx} \times N_{rx})} \sim \mathcal{N}(\mu, \sigma^2)}{\sqrt{2}} + \frac{j\mathbf{H}_{2(N_{tx} \times N_{rx})} \sim \mathcal{N}(\mu, \sigma^2)}{\sqrt{2}} \quad (9)$$

donde **H'** es la matriz H estimada, compuesta por N_H matrices de tamaño $N_{tx} \times N_{rx}$, sumadas M veces, que corresponde con el número de muestras utilizadas. La capacidad final será la media entre todas las muestras obtenidas.

IV. PRUEBAS Y OBTENCIÓN RESULTADOS

Para simular el sistema completo se han utilizado herramientas de simulación de desarrollo propio basadas

en eventos discretos. El sistema completo incluye desde los nodos móviles hasta la red de acceso, teniendo en cuenta el RAN y la red de *backhaul*, así como la interacción de todos los elementos. El proceso de evaluación de los KPIs se realiza en cada cambio de posición del usuario, que sucede cuando su modelo de movimiento genera un cambio (en función de la velocidad y distancia recorrida). En este evento las condiciones del canal cambian, por lo que se realizan los cálculos de los diferentes parámetros.

Las estaciones base están colocadas según una distribución de Poisson (PPP) por todo el escenario. Los UE se mueven por el escenario siguiendo un modelo de movimiento basado en distribuciones aleatorias (*Random WayPoint*), y el tipo de usuario está repartido de la siguiente manera:

- 100 usuarios de tipo eMBB, que generan un tráfico de tipo vídeo VOD, FTP, AR/VR y web.
- 100 usuarios de tipo URLLC, que generan un tráfico de tipo V2X y gaming.
- 200 usuarios de tipo mMTC, que generan tráfico de tipo web, IoT y FTP.

La red de acceso está formada por 4 routers de borde, cada cual está conectado a una estación base de tipo macro. Los enlaces entre los routers siguen una malla completa. Estos enlaces están dimensionados con una capacidad de 10 Gbps.

Para comprobar el rendimiento de los mecanismos propuestos, se ha realizado una batería de simulaciones para la obtención de resultados, y el posterior tratamiento de los mismos. Para minimizar la aleatoriedad de el posicionamiento de las BS y el movimiento de los usuarios, se han realizado 35 simulaciones con la misma configuración para cada una de las pruebas, cada una de ellas repetida 3 veces: una con cada planificación de enlace; en total, 105 ejecuciones. Todos los resultados, por tanto, son valores medios de todas las simulaciones. La configuración de los modelos para la ejecución se muestra en la Tabla II.

Parámetro	Macro BS	Small BS	UE
Nº	4	20	400
Altura	40 m	15 m	1.7 m
Nº antenas MIMO	10	10	4
Frecuencia	6 GHz	21 GHz	Banda uplink
Ganancia	12 dBi	8 dBi	5 dBi
Potencia de transmisión	28 dBm	12 dBm	5 dBm
Ancho de banda	300 MHz	900 MHz	-
Tiempo de *slot* (piloto)	10 ns	10 ns	10 ns
Tiempo de simulación			1 hora
Tamaño del escenario			144 Km^2 (UMa)
Capacidad media de los enlaces de backhaul			1 Gbps

Tabla II
PARÁMETROS UTILIZADOS EN LA SIMULACIÓN.

En la Figura 2 se muestran los resultados obtenidos que corresponden al número de conexiones bloqueadas por la red para cada tipo de usuario. Como el tráfico se genera según distribuciones estocásticas, se considera el porcentaje con respecto al total. Se considera que una conexión sufre bloqueo (y su posterior descarte) cuando el enlace (bien sea entre el UE y la BS o en la red de backhaul) no tiene capacidad suficiente para satisfacer la demanda o cuando su latencia media es mayor a la requerida por el tráfico que está pasando por el enlace. Bajo las mismas condiciones, se modifican las planificaciones de enlace para comprobar la eficacia que tiene cada una de ellas con el tipo de usuario al que van dirigido.

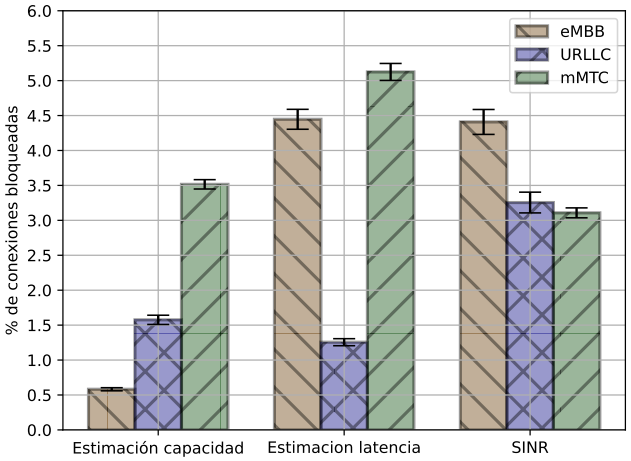


Fig. 2. Porcentaje de conexiones bloqueadas con respecto al total, clasificadas por tipo de usuario según requerimientos (capacidad, latencia o densidad de conexión), en función del tipo de planificación de enlace.

Para observar de manera absoluta la evolución de las KPIs que se pretenden mejorar, se presentan los resultados de capacidad y latencia medias según los distintos mecanismos desarrollados a lo largo del tiempo de las simulaciones. Las líneas de error corresponden a unos intervalos de confianza del 95%, calculados con las 35 simulaciones realizadas.

En la Figura 3 observamos la capacidad máxima teórica disponible, medida en Mbps, por los enlaces entre los usuarios y las estaciones base a las que están conectados (en valor medio) según las planificaciones de enlace desarrolladas de estimación de latencia y capacidad, y la planificación de referencia por SINR.

Se observa un aumento de la capacidad si utilizamos la estimación del canal para conectar a los usuarios a las estaciones base, y no únicamente el nivel de señal recibida. Se aumenta la capacidad media de los enlaces en un 21'9% si se utiliza el estimador de capacidad para estimar la capacidad máxima. El sombreado de las gráficas corresponde a los intervalos de confianza de las 35 ejecuciones realizadas. Los intervalos de confianza están ajustados al 95%.

En la Figura 4 se muestra, de la misma manera, la latencia de los enlaces que conectan a los usuarios con el RAN, medida en milisegundos, con los mecanismos propuestos, tomando como referencia la planificación por nivel de señal.

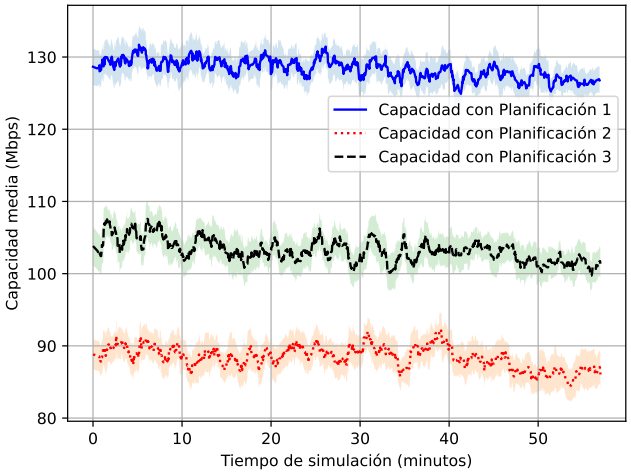


Fig. 3. Capacidad disponible por los usuarios de la simulación en función de la planificación de enlace. Planificación 1: estimación de capacidad; planificación 2: estimación de latencia; planificación 3: planificación de referencia basada en SINR.

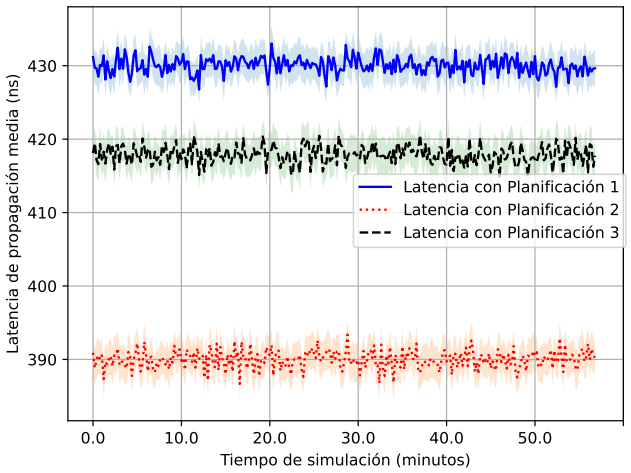


Fig. 4. Latencia de los enlaces entre los usuarios y las estaciones base a las que están conectados en función de la planificación de enlace. Planificación 1: estimación de capacidad; planificación 2: estimación de latencia; planificación 3: planificación de referencia basada en SINR.

Se observa, además de una disminución del 7% en la latencia media de los enlaces, una estabilización en la latencia de las comunicaciones, que puede ser utilizada para una mayor sincronización de las conexiones. Los intervalos de confianza están ajustados al 95%.

V. CONCLUSIONES

En este trabajo se presentan dos mecanismos basados en estimación del canal para las decisiones de planificación de enlace de los terminales móviles de las redes celulares. Estos mecanismos son aplicados a la modelización de un RAN con tecnología 5G y se centran en la evaluación de dos parámetros importantes: la capacidad y la latencia. Las pruebas se han realizado en herramientas de simulación de desarrollo propio basadas en eventos discretos, que evalúan los parámetros cuando cambian las condiciones del canal.

Con los resultados obtenidos de las pruebas realizadas se pueden obtener las siguientes conclusiones:

La planificación de enlace propuesta basada en estimación de latencia minimiza el número de conexiones bloqueadas para usuarios de tipo URLLC. Esto mejora la fiabilidad de las conexiones críticas como el caso de los vehículos autónomos, y dota la red de mayor robustez para conexiones de baja latencia, reduciéndose en un 7%. Del mismo modo que la planificación de enlace basada en estimación de capacidad aumenta la capacidad media disponible en los enlaces de la generalidad de los usuarios en un 21%, factor clave para reducir el número de conexiones bloqueadas en los usuarios de tipo eMBB. Para los usuarios de tipo mMTC no se observan resultados concluyentes.

La estimación del canal permite el desarrollo de mecanismos de nivel físico que mejoren el rendimiento del RAN o las capas superiores, minimizando el número de conexiones bloqueadas, descongestionando los enlaces de la red y aumentando, por tanto, la capacidad de las redes celulares. Descentralizando el costo del envío de pilotos y haciendo recaer el cómputo sobre los nodos móviles, se reduce la complejidad de las estaciones base. Estas sólo tendrán que emitir pilotos cuando un usuario necesite menor latencia o mayor capacidad, y no de manera periódica como en los mecanismos convencionales.

Sin necesidad de mejorar la tecnología de las redes existentes, es posible optimizar el uso de los recursos mediante mecanismos que mezclan soluciones software con tratamiento de señal. Estas soluciones no plantean un gran cambio en la infraestructura, y permiten gran escalabilidad. El uso de señales pilotos permite, por tanto, distinguir desde el punto de vista de la red, las necesidades de los usuarios sin necesidad de un establecimiento de la conexión.

Como trabajos futuros podría destacarse la aplicación de estos mecanismos para la reserva de recursos en una red con tecnología *RAN-Slicing*. Las estaciones base comparten una serie de recursos (capacidad, latencia, etcétera) que serán utilizados por una misma *slices*. Los usuarios conectados a esa *slice* tendrán todos esos recursos accesibles independientemente del punto de acceso a la red al que estén conectados. Estos mecanismos pueden ayudar al orquestador de las *slices* a reservar recursos para poder aumentar el rendimiento de los enlaces y minimizar la pérdida de conexiones.

AGRADECIMIENTOS

Este trabajo ha sido financiado parcialmente por la Consejería de Economía e Infraestructuras de la Junta de Extremadura con el proyecto IB18003 y la Ayuda GR18141, y por el Ministerio de Ciencia, Innovación y Universidades con el proyecto RTI2018-102002-A-I00.

REFERENCIAS

[1] *Digital 2021: Global Overview Report*. Accessed: May. 10, 2021. [Online]. Available: https://datareportal.com/reports/digital-2021-global-overview-report

[2] D. Reinsel, J. Gantz and J. Rydning, "The Digitization of the World from Edge to Core", Seagate Technology, California, USA. White paper, 2018.

[3] P. Popovski, K. F. Trillingsgaard, O. Simeone and G. Durisi, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," in *IEEE Access*, vol. 6, pp. 55765-55779, 2018.

[4] G. Pocovi, K. I. Pedersen and P. Mogensen, "Joint Link Adaptation and Scheduling for 5G Ultra-Reliable Low-Latency Communications," in *IEEE Access*, vol. 6, pp. 28912-28922, 2018.

[5] Z. Naghsh and S. Valaee, "Conflict-Free Scheduling in Cellular V2X Communications," in *IEEE/ACM Transactions on Networking*, vol. 29, no. 1, pp. 106-119, Feb. 2021.

[6] T. E. Bogale, X. Wang and L. B. Le, "Adaptive Channel Prediction, Beamforming and Scheduling Design for 5G V2I Network: Analytical and Machine Learning Approaches," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5055-5067, May 2020.

[7] H. D. R. Albonda and J. Pérez-Romero, "An Efficient RAN Slicing Strategy for a Heterogeneous Network With eMBB and V2X Services," in *IEEE Access*, vol. 7, pp. 44771-44782, 2019.

[8] T. Guo and A. Suárez, "Enabling 5G RAN Slicing With EDF Slice Scheduling," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2865-2877, March 2019.

[9] A. Ramírez-Arroyo, P. H. Zapata-Cano, Á. Palomares-Caballero, J. Carmona-Murillo, F. Luna-Valero & J. F. Valenzuela-Valdés, "Multilayer Network Optimization for 5G & 6G," in *IEEE Access*, vol. 8, pp. 204295-204308, 2020.

[10] 3GPP TR 36.873, V12.1.0, "Study on 3D channel model for LTE (release 12)," Mar. 2015.

[11] S. Sun *et al.*, "Investigation of Prediction Accuracy, Sensitivity, and Parameter Stability of Large-Scale Propagation Path Loss Models for 5G Wireless Communications," in *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2843-2860, May 2016, doi: 10.1109/TVT.2016.2543139.

[12] Foschini, G. & Gans, M. "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas", in *Wireless Personal Communications* 6, 311–335 (1998).

[13] D. A. Chekired, M. A. Togou, L. Khoukhi & A. Ksentini, "5G-Slicing-Enabled Scalable SDN Core Network: Toward an Ultra-Low Latency of Autonomous Driving Service," in *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 8, pp. 1769-1782, 2019.

[14] Rel-15, document TR 21.915, 3GPP, 2018.

[15] M. Gapeyenko et al., "On the Temporal Effects of Mobile Blockers in Urban Millimeter-Wave Cellular Scenarios," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10124-10138, 2017.

[16] S. Pratschner, S. Schwarz and M. Rupp, "Single-user and multi-user MIMO channel estimation for LTE-Advanced uplink," 2017 *IEEE International Conference on Communications (ICC)*, 2017, pp. 1-6.

[17] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz and J. M. Lopez-Soler, "A Survey on 5G Usage Scenarios and Traffic Models," in *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 905-929, Secondquarter 2020.

[18] "3rd Generation Partnership Project; technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)," Tech. Rep. 3GPP TR 36.814 V9.2.0, Mar. 2017.

[19] Carrascosa, Marc & Bellalta, Boris. (2020). Cloud-gaming:Analysis of Google Stadia traffic. arXiv:2009.09786.

[20] X. Wang, T. Kwon, Y. Choi, M. Chen and Y. Zhang, "Characterizing the gaming traffic of World of Warcraft: From game scenarios to network access technologies," in *IEEE Network*, vol. 26, no. 1, pp. 27-34, January-February 2012.

[21] "LTE physical layer framework for performance verification," Tech. Rep. 3GPP TSG-RAN1#48 R1-070674, Feb. 2007.

[22] Y. Al Mtawa, A. Haque and B. Bitar, "The Mammoth Internet: Are We Ready?," in *IEEE Access*, vol. 7, pp. 132894-132908, 2019.

[23] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on RAN Improvements for Machine-type Communications (Release 11)," Tech. Rep. 3GPP TR 37.868 V11.0.0, Sep. 2011.

[24] F. Hu, Y. Deng, W. Saad, M. Bennis and A. H. Aghvami, "Cellular-Connected Wireless Virtual Reality: Requirements, Challenges, and Solutions," in *IEEE Communications Magazine*, vol. 58, no. 5, pp. 105-111, May 2020.

[25] S. Friston et al., "Quality of Service Impact on Edge Physics Simulations for VR," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2691-2701, May 2021.

[26] Wang, J., Shao, Y., Ge, Y., & Yu, R. (2019). A survey of vehicle to everything (V2X) testing. *Sensors*, 19(2), 334.

[27] Storck CR, Duarte-Figueiredo F. A 5G V2X Ecosystem Providing Internet of Vehicles. *Sensors*. 2019; 19(3):550.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Generación de escenarios de propagación mediante modelos generativos y aprendizaje por refuerzo

Natalia Mártir-Moreno, Alejandro Ramírez-Arroyo, Sohrab Vafa, Luz García, y Juan F. Valenzuela-Valdés
Departamento de Teoria de la Señal, Telemática y Comunicaciones
Universidad de Granada
nataliamartir@ugr.es, alera@ugr.es, sohrabvafa@correo.ugr.es, luzgm@ugr.es, juanvalenzuela@ugr.es.

Abstract: Las comunicaciones 6G esperan generar nuevos escenarios en un mundo completamente interconectado. Si las comunicaciones 5G se centran en proporcionar infraestructura celular terrestre, las redes 6G van un paso más allá proporcionando cobertura global integrada para comunicaciones por satélite y drones, comunicaciones terrestres, marítimas y submarinas. Por primera vez, estas redes se conciben de manera que la Inteligencia Artificial sea una de las características clave de las nuevas arquitecturas. Para desarrollar redes eficientes e inteligentes, es necesario caracterizar, clasificar y generar escenarios de comunicaciones futuras para su estudio. A partir de las medidas adquiridas en canales de propagación controlados, se propone generar nuevos escenarios de comunicaciones mediante el uso de Aprendizaje por Refuerzo.

***Palabras Clave*- Comunicaciones inalámbricas, Inteligencia Artificial, clustering, VAEs, propagación.**

I. INTRODUCCIÓN

Hoy en día, los avances en las redes de comunicaciones han dado como resultado un mundo completamente conectado. La Inteligencia Artificial (IA) se aplica ya en todas las capas/niveles de protocolos de las redes 5G. Además, la combinación de IA y técnicas de computación está siendo explorada en el campo de la investigación. Por lo tanto, las aproximaciones basadas en IA serán uno de los pilares fundamentales para que las redes 6G cumplan con los requisitos de millones de usuarios en todo el mundo [1][2]. Se requieren redes inteligentes 6G para negociar un equilibrio entre el rendimiento de la red, la respuesta instantánea y el consumo de energía, para lo que una configuración óptima de la potencia de transmisión en las estaciones base logra un equilibrio energético en la red [3]. Para alcanzar este equilibrio es necesario realizar optimizaciones multiobjetivo para cada uno de los diferentes nuevos escenarios 6G como los que se muestran en la Figura 1.

En este marco de futuras comunicaciones se definen los siguientes tres escenarios tipo en función de los requisitos del usuario:

- ***further-enhanced Mobile BroadBand (further-eMBB).*** Este esquema tiene como objetivo proporcionar al usuario velocidades de datos en el rango de TB/s.
- ***enhanced ultra–Reliable Low Latency Communications (enhanced uRLLC).*** Este enfoque se centra en minimizar el retraso de un extremo a otro en las redes de comunicación. Para ello, nuevos esquemas de red menos complejos en su arquitectura permitirán un procesamiento más rápido de los paquetes de datos.
- ***ultra-massive Machine-Type Communications (ultra-mMTC).*** Basado en el paradigma de Internet de las cosas (IoT), este enfoque propone un esquema de red donde múltiples dispositivos están completamente interconectados.

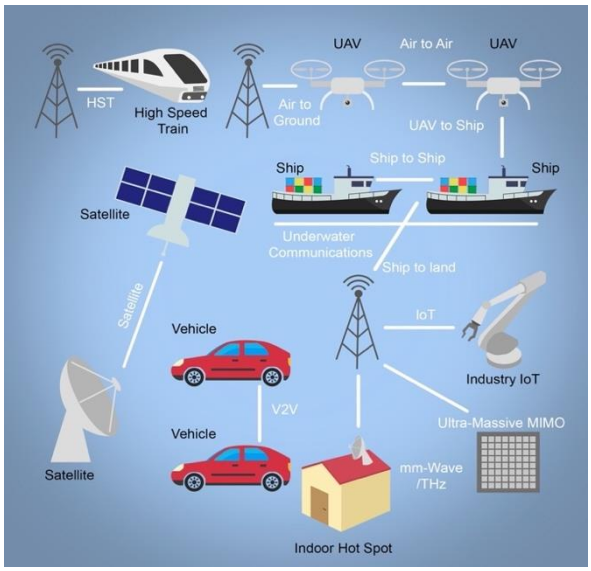


Fig. 1. Nuevos escenarios para las futuras generaciones de redes móviles.

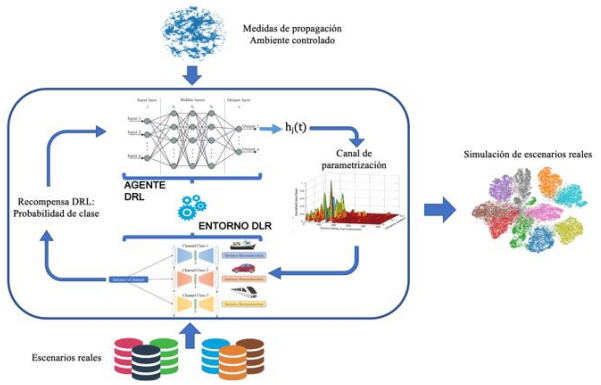
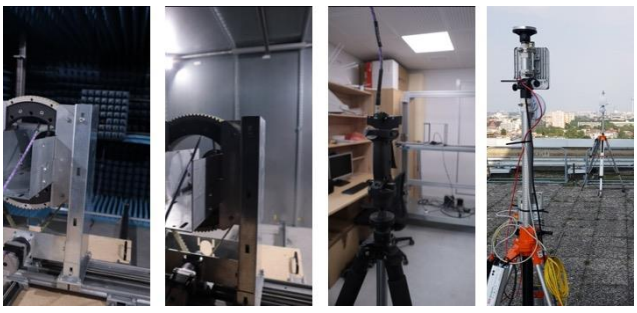


Fig. 2. Esquema del enfoque de Deep Learning para la generación de nuevos escenarios de tráfico real.

Todos estos avances tendrán que hacerse con nuevas arquitecturas de red que permitan una cobertura global de las redes, por lo que 6G debe proporcionar una red integrada Espacio-Aire-Tierra-Mar [4]. Por lo tanto, en las redes 6G aparecerán nuevos escenarios que deberán caracterizarse. En estos nuevos escenarios, las comunicaciones Ship-to-Ship (S2S), Vehicle-to-Satellite (V2S) o UAV-to-UAV, solo se pueden desarrollar de manera fiable si tenemos una caracterización previa de los entornos de red.

Resumiendo lo anterior, es necesario un conocimiento y modelado del canal de comunicaciones para analizar y optimizar su impacto físico en las señales radio de comunicación. Con ello se consigue diseñar tecnologías de comunicación efectivas y viables que respondan a las necesidades del usuario. Actualmente, se están comenzando a desarrollar técnicas de IA para estimar el canal de comunicación. Una de las propuestas de este artículo es dar un paso más, no solo para estimar el canal de comunicación a través de la IA sino para clasificar y generar futuros canales de comunicaciones. En particular, dados los nuevos y desafiantes esquemas de redes futuras que coexisten con entornos de medición controlados, como las cámaras anecoicas y reverberantes, la IA puede desempeñar un papel importante para emular los casos de tráfico real trabajando con grandes cantidades de medidas de canal obtenidas de un entorno controlado. Al centrarnos en la parte de IA, relacionada con los paradigmas de aprendizaje automático, el objetivo es aprender a combinar y procesar mediciones controladas para crear escenarios de tráfico



(a) (b) (c) (d)

Fig. 3. Imágenes de los cuatro escenarios ilustrados en la prueba de concepto: (a) Anecoico, (b) Reverberante, (c) Interior y (d) Exterior. (a), (b) y (c) han sido adquiridos en las instalaciones del grupo SWAT de la Universidad de Granada. (d) se ha medido en el Fraunhofer Heinrich-Hertz-Institut [6].

real deseados sin tener un modelo exacto. La aportación de este trabajo es la propuesta y prueba de concepto de un nuevo banco de pruebas y metodología de trabajo basada en IA para clasificar y generar escenarios de propagación.

II. Generación de escenarios mediante aprendizaje por refuerzo (Deep Reinforcement Learning)

El aprendizaje por refuerzo imita el proceso cognitivo del cerebro humano explorando acciones de aprendizaje por prueba y error en los casos en los que no se puede aplicar una estrategia estructurada de aprendizaje automático supervisado. Se define un agente que interactúa con un entorno para explorar las mejores acciones para representar indirectamente un modelo, maximizando una cierta recompensa a largo plazo. El aprendizaje por refuerzo se propone para una amplia gama de aplicaciones de futuras comunicaciones inalámbricas, funcionando especialmente bien cuando se utilizan arquitecturas de redes profundas.

El aprendizaje por refuerzo profundo asegura la convergencia de los algoritmos de aprendizaje con un coste inferior en términos de tiempo comparado con aprendizaje por refuerzo clásico [5].

A. Procesamiento mediante DRL

Nuestro trabajo propone aplicar DRL para procesar y combinar un conjunto de respuestas en frecuencia y en tiempo de un canal de comunicaciones, con el objetivo de crear nuevas respuestas complejas de diversos escenarios de tráfico real. La Figura 2 muestra un esquema de este enfoque:

- El **agente DRL** se implementa con una red neuronal profunda que aprende a combinar un conjunto de respuestas de canal controladas originadas en el banco de pruebas. Como resultado, genera instancias de la respuesta del canal emulado.
- El **entorno DRL** se implementa utilizando modelos generativos profundos en competencia, para la clasificación de los escenarios emulados. Cada modelo se entrena con una amplia colección de datos de cada escenario real. Los modelos generativos como los *Variational Autoencoders* (VAEs) o las Redes Generativas Antagónicas (*Generative Adversarial Networks*, GANs) aprenden la verdadera distribución de los datos de entrenamiento para generar nuevas instancias de datos con algunas variaciones.
- La **recompensa DRL** se obtiene cuando la respuesta de canal emulada por el agente DRL, es clasificada por el entorno DRL. La probabilidad de que el canal emulado pertenezca a la clase objetivo de la emulación será la recompensa. Si se quiere emular uno de los escenarios existentes, la recompensa será proporcional a la probabilidad de que el canal emulado pertenezca a la clase del escenario en particular.

El ciclo Emulación de canal - Clasificación de canal se repite de forma recursiva hasta la convergencia del algoritmo DRL. Como prueba de concepto, el resto de esta sección proporciona al lector un ejemplo de clusterización,

clasificación y emulación de cuatro escenarios diferentes a través de DRL. Estos escenarios se miden en la banda de 24.25-27.5 GHz, específicamente reservada para la asignación de espectro 5G en Europa. Además, incluimos medidas en la banda de ondas milimétricas (60 GHz) [6] que serán fundamentales en las futuras comunicaciones móviles.

B. Propiedades del canal

Para aplicar DRL es fundamental comprender y extraer las propiedades que hacen que cada tipo de escenario sea único. El canal se puede analizar tanto en el dominio del tiempo como en el de la frecuencia y de cada uno se pueden extraer diferentes propiedades discriminatorias. Por un lado, en el dominio del tiempo adquirimos los relacionados con el tiempo de llegada de las componentes multicamino del canal. Por otro lado, en el dominio de la frecuencia obtenemos métricas relacionadas con la atenuación y fase del canal. Por ejemplo, algunas de estas propiedades son:

- **Factor K**: Es la relación entre la componente principal de los *MultiPath Components* (MPCs) y la potencia total del resto de MPCs en el escenario.
- **Retardo promedio** (τ_{mean}), **retardo de varianza** (τ_{var}) y **dispersión del retardo** (τ_{RMS}): Estas tres propiedades están relacionadas con el tiempo de llegada de la señal. τ_{mean} indica el tiempo promedio en que la potencia de la señal recorre el par TX-RX. τ_{var} y τ_{RMS} son indicativos de la separación temporal de las distintas componentes multicamino.
- **Atenuación:** En el dominio de la frecuencia, esta propiedad representa la atenuación de la onda electromagnética debido a la propagación en el escenario.
- **Eficiencia espectral:** Determina la cantidad máxima de información (en bps/s/Hz) que se puede transmitir de manera fiable a través del canal de comunicaciones.

A modo de ejemplo, mostramos cómo esta parametrización discriminativa de canales podría funcionar en cuatro escenarios diferentes. Los escenarios elegidos son una cámara anecoica, una cámara reverberante, un escenario *indoor* y un escenario *outdoor* (ver Figura 3). Se adquieren 1089 medidas para cada escenario, para un total de 4356 canales. Los dos primeros escenarios pertenecen a la familia de escenarios controlados. Los dos últimos pertenecen a la familia de escenarios reales que podrían ocurrir en un entorno real de comunicaciones móviles. Los tres primeros escenarios se miden en la banda de 26 GHz, mientras que el último se adquiere en la banda de 60 GHz [6].

C. Clusterización, clasificación y generación de escenarios.

Como se dijo anteriormente, cada escenario se puede caracterizar por un conjunto de parámetros discriminativos basados en sus propiedades. En el ejemplo proporcionado, se han utilizado el Factor K, τ_{mean}, τ_{var}, τ_{RMS}, la atenuación y la eficiencia espectral. Dado que estas seis propiedades producen un espacio de seis dimensiones que no es fácilmente interpretable a primera vista, estamos interesados en reducir la dimensionalidad a un espacio bidimensional. Específicamente, *t-Distributed Stochastic Neighbor Embedding* (t-SNE) [7] es una técnica no lineal para la reducción de dimensionalidad particularmente adecuada para la visualización en un espacio bidimensional de conjuntos de datos de mayor dimensión. Al aplicar t-SNE a las mediciones de nuestros escenarios, podemos visualizar el grado de clusterización de estos cuatro escenarios como se muestra en la Figura 4(a). t-SNE aprovecha las características discriminativas de cada escenario para formar cuatro grandes grupos. Una vez que t-SNE detecta las peculiaridades de cada escenario, puede distinguir cuatro grupos principales, uno para cada escenario.

Una vez definimos las técnicas t-SNE para la reducción de la dimensionalidad, se proceden a usar modelos generativos, en este caso los VAEs.

En la prueba de concepto presentada, hemos seleccionado el VAE [8] como modelo generativo para modelar las verdaderas distribuciones de datos de los escenarios. Los VAE son arquitecturas de aprendizaje automático no supervisadas formadas por dos redes neuronales, denominadas codificador y decodificador respectivamente. El objetivo del codificador es disminuir la dimensión de las entradas al VAE, creando un conjunto de parámetros en un espacio latente. El decodificador es responsable de regenerar los datos originales a partir de dichos parámetros en el espacio latente.

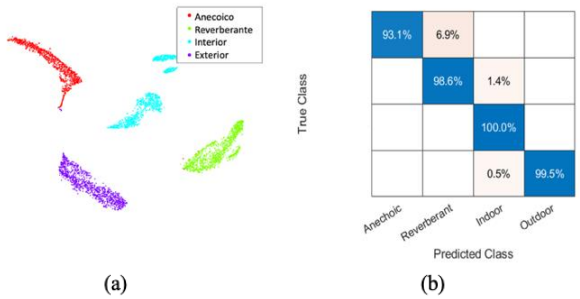


Fig.4. (a) Agrupación de cuatro escenarios mediante la aplicación de t-SNE y (b) clasificación mediante el uso de VAEs.

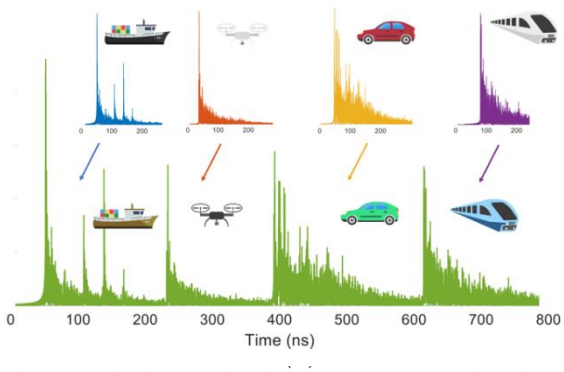


Fig 5. Ejemplo de generación de nuevos escenarios combinando cuatro escenarios reales. Cada canal se transforma para cambiar sus propiedades físicas. Tenga en cuenta que el cambio de color en el icono entre el escenario original y el generado representa la manipulación física del escenario.

En nuestro ejemplo se entrenan de manera no supervisada 4 VAEs, usando en cada uno de ellos datos de cada uno de los cuatro escenarios que se quieren modelar. Una vez entrenados los VAEs para cada uno de los 4 escenarios, el nuevo escenario emulado que se quiere clasificar se usa para regenerar datos en los 4 VAEs. Se elegirá como clase de pertenencia del canal emulado, la clase del VAE que genere muestras con una entropía cruzada más baja. La Figura 4(b) muestra un ejemplo simple de clasificación de los cuatro escenarios siguiendo la estrategia descrita. Se puede ver que la clasificación se produce de manera satisfactoria, con precisiones superiores al 93% en todas las clases.

En el paso final del proceso de aprendizaje, es viable generar el canal en base a las recompensas que brinda el entorno DRL. El agente DRL elige aquellos canales que, agregados linealmente, recrean un escenario dado con mayor precisión. Además, el agente no solo elige los mejores canales, sino que también incluye un retardo y atenuación para cada escenario. Esta técnica de postprocesamiento implica un retraso en el canal o una disminución en la potencia de transmisión. La Figura 5 muestra un ejemplo de generación de nuevos escenarios. Este ejemplo se genera modificando cuatro escenarios reales, cada uno representado en un color diferente. Cada uno de los nuevos escenarios se combina de forma lineal agregando una determinada atenuación y *delay* de manera que la nueva respuesta al impulso es representativa de un escenario completamente nuevo.

III. Conclusiones

Este trabajo propone la parametrización, clusterización y clasificación de escenarios 6G basados en Aprendizaje por Refuerzo Profundo y Modelos Generativos. Utilizando un conjunto de medidas de entorno controlado y escenarios reales como entrada, esta idea define un agente de aprendizaje que emula nuevos escenarios a través de una red neuronal profunda clásica. Las recreaciones son recompensadas con el resultado de un entorno de clasificación implementado como un conjunto de autoencoders variacionales (VAEs) entrenados con escenarios reales. Estos modelos generativos utilizados para la clasificación están orientados a conocer las verdaderas distribuciones de datos de los escenarios a clasificar y, por tanto, proporcionan herramientas útiles para modificar las distribuciones de datos de acuerdo con las necesidades potenciales.

Como líneas futuras, el uso de redes generativas antagónicas para combinar escenarios bajo demanda siguiendo las estrategias exitosas aplicadas en otros campos del conocimiento pueden abrir nuevas líneas de investigación. La generación de escenarios presentada ayudará a modelar y analizar en profundidad los nuevos canales de comunicaciones móviles. Estos modelos permitirán el estudio y mejora de las prestaciones, por ejemplo en ámbitos tan demandados como la eficiencia energética.

Agradecimientos

Este trabajo ha sido financiado por el Programa Español de Investigación, Desarrollo e Innovación bajo el Proyectos RTI2018-102002-A-I00, el Proyecto PID2020-112545RB-C54, y el Proyecto TIN2016-75097-P, y en parte por la “Junta de Andalucía” bajo el Proyecto B-TIC-402-UGR18, el Proyecto A-TIC-608-UGR20, y Proyecto P18.RT. 4830 y en parte por la beca predoctoral FPU19/01251. Los autores agradecen a Carmelo García-García, Ángel Palomares-Caballero, Carlos Molero-Jiménez por sus ayuda y comentarios, y a José Francisco Cortés-Gómez por el apoyo gráfico.

Referencias

[1] C. -X. Wang, J. Huang, H. Wang, X. Gao, X. You and Y. Hao, "6G Wireless Channel Measurements and Models: Trends and Challenges," IEEE Vehicular Technology Magazine, vol. 15, no. 4, pp. 22-32, Dec. 2020.

[2] Z. Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28-41, Sept. 2019.

[3] A. Ramírez-Arroyo, P. H. Zapata-Cano, Á. Palomares-Caballero, J. Carmona-Murillo, F. Luna-Valero and J. F. Valenzuela-Valdés, "Multilayer Network Optimization for 5G & 6G," IEEE Access, vol. 8, pp. 204295-204308, 2020.

[4] I. F. Akyildiz, A. Kak and S. Nie, "6G and Beyond: The Future of Wireless Communications Systems," IEEE Access, vol. 8, pp. 133995- 134030, 2020.

[5] N.C. Loung, D.T. Hoang, D. Niyato, P. Wang, Y.C. Liang, D. In Kim, “Applications of Deep Reinforcement Learning in Communications and Networking: A Survey”, IEEE Communications Surveys and Tutorials, vol. 21, no 4, 2019.

[6] mmWave Channel Model Alliance Database, Fraunhofer Heinrich-Hertz-Institut, Berlin, Germany. [Online]. Available: https://5gmm.nist.gov/

[7] L.J.P. van der Maaten and G.E. Hinton, “Visualizing High-Dimensional Data Using t-SNE,” Journal of Machine Learning Research vol. 9, pp. 2579-2605, 2008.

[8] D. Kingma and M. Welling, “Auto-Encoding Variational Bayes” in 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, 2014.

[9] A. Jabbar, X. Li and B. Omar, “A Survey on Generative Adversarial Networks: variants, applications and training”, 2020, doi:arXiv:1909.11573


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Retardo en redes fronthaul con split funcional flexible: un modelo basado en teoría de colas

Luis Diez, Ramón Agüero
Departamento de Ingeniería de Comunicaciones. Universidad de Cantabria.
{ldiez, ramon}@tlmat.unican.es

En este trabajo se estudia el retardo en topologías de red vRAN, considerando tanto las estaciones base, que se dividen entre un controlador y un cabezal de radio remoto, y la red de conmutación de paquetes fronthaul que los une. Se contempla el uso de functional split flexible, según el que las funciones que se ejecutan en cada una de las dos entidades se puede modificar dinámicamente. Se propone un modelo basado en teoría de colas, que es capaz de reflejar de manera precisa el comportamiento de estos nodos, que se valida tras una extensa campaña de medidas. Además, se utiliza la teoría de redes abiertas de Jackson para modelar el retardo extremo a extremo en la red fronthaul, lo que permite analizar el impacto de establecer diferentes políticas de red. Los resultados ponen de manifiesto que el modelo propuesto se puede emplear para establecer las configuraciones óptimas de red, pues los resultados que ofrece son prácticamente idénticos a los obtenidos mediante simulación.

***Index Terms*—vRAN, functional split, teoría de colas, redes de Jackson**

I. Introducción

Uno de los requisitos más exigentes en los sistemas 5G es el relativo al retardo, que resulta fundamental para soportar adecuadamente servicios de tipo *Ultra-Reliable Low Latency Communication* (URLLC), tales como los relativos a Internet táctil o a la conducción autónoma. Por otro lado, las arquitecturas de redes de acceso radio están sufriendo una continua evolución, incorporando, entre otros, elementos SDN y NFV, dando lugar a lo que ya se conoce como virtual RAN. Aunque la capacidad de virtualización de funciones de las estaciones base tiene grandes ventajas (como reducción de costes), también aparecen nuevos aspectos que deben ser analizados, tales como el retardo asociado a esta virtualización.

Inicialmente, las soluciones Cloud-RAN (C-RAN) proponían la completa virtualización de las unidades de banda base, con una conexión de gran capacidad con los antenas. En este artículo se considera una evolución de este solución, en la que existen diferentes grados de centralización (*functional-splits*) [1] y esta centralización puede ser modificada de manera dinámica, lo que da lugar a la funcionalidad conocida como *flexible functional-split*. Esta adaptación permite solventar algunas de las limitaciones de las soluciones C-RAN [2] iniciales, adaptando la red a las necesidades concretas y recursos que se tienen. En estas arquitecturas la estación base se divide en una *centralized unit* (CU) que contiene ciertas funciones y que se conecta, a través de una red de conmutación de paquetes *fronthaul*, a las *distributed units* (DU) en las que están el resto de funciones. A su vez las DUs poseen una conexión de gran capacidad con las antenas o *radio units*. A fin de soportar servicios que precisen URLLC, es necesario conocer el retardo asociado entre las CUs y DUs. En este artículo se extiende el modelo presentado inicialmente en [3], que permite analizar el retardo asociado a una CU o DU, para conocer la latencia extremo a extremo en la red *fronthaul* cuando se aplica una determinada política de selección de split. El modelo propuesto podría ayudar al dimensionado de este tipo de redes, y a establecer límites en el tráfico admisible ante ciertos requisitos de retardo.

El resto del documento se estructura de la siguiente manera. En la Sección II se presenta una revisión de la literatura relativa a *flexible functional-split*, resaltando el carácter innovador del modelo propuesto. A continuación, en la Sección III se presenta el modelo basado en cadenas de *Markov* y teoría de redes de *Jackson*, que se valida en la Sección IV sobre diferentes escenarios. en la Sección V se presentan las conclusiones más relevantes que se han obtenido, y se enumeran líneas de trabajo futuro.

II. Trabajos relacionados

El potencial de las arquitecturas *functional-split* flexibles se ha analizado ampliamente en la literatura [4], [3], y ya existen trabajos describiendo su implementación para posibilitar la selección dinámica del nivel de centralización, tales como [5], [6].

Más relacionados con este trabajo, han aparecido propuestas de políticas de selección de *split* centradas en diferentes aspectos. Por ejemplo, en [7], [8] Harutyunyan *et al.* modelan la selección de split como un problema de tipo *Virtual Network Embedding* (VNE) formulado como

un *Integer Linear Program* (ILP). De forma similar, los autores de[9] y [10] proponen algoritmos de selección de *split* que asegure el uso de técnicas de cooperación entre elementos de acceso, mientras se hace un uso eficiente de la red *fronthaul*, permitiendo el despliegue de servicios que requieran URLLC.

Otros trabajos prestan atención a métricas diversas en sus políticas de selección de *split*, tales como la tasa [11], o el retardo [12]. Entre los parámetros considerados, existen multitud de propuestas que se centran en la eficiencia energética, tales como [13], [14], [15]. Otro grupo de trabajos se centran en la interacción de la selección de *split* con la red *fronthaul* óptica. En este sentido se ha analizado tanto la reducción de latencia [16] como la limitación de capacidad de la red [17]. Asimismo, existen trabajos que presentan soluciones de *orquestación* [18], que permitan la reconfiguración global de la red de acceso ante cambios en el nivel de centralización.

Aunque la revisión de la literatura se podría extender, la mayoría de las investigaciones previas proponen soluciones para definir el nivel de centralización. Por el contrario, el modelo presentado en este trabajo tienen como objetivo modelar el comportamiento de la red de *fronthaul*, en función del retardo, cuando se aplica cualquier política.

III. Modelo de colas para el *fronthaul*

En esta sección se va a presentar el modelo, basado en teoría de colas, que considera dos tipos de nodos: (1) refleja el comportamiento del CU o DU; y (2) se usan para modelar *switches* y enlaces en la red. El segundo tipo se modelará mediante un nodo M/M/1, mientras que el primero precisa una aproximación más compleja. A continuación se describirá el modelo de los nodos que representan los CU y DU, para posteriormente establecer el retardo extremo a extremo esperado en la red de *fronthaul*. Para facilitar la lectura, la Tabla I enumera los símbolos que se utilizan en el resto de la sección.

Tabla I: Símbolos y variables

	Nodos CU/DU
s	Número de *splits*
λ	Tasa de llegada de tramas
μ_j	Tasa de servicios del *split* j^{th}
$\alpha_{j,k}$	Probabilidad de transitar del *split* j^{th} al k^{th} $\sum_{k=1}^{s}\alpha_{j,k}=1,\ \alpha_{j,j}=0$
γ_j	Tasa de cambio del *split* j^{th}
ξ_j	Inverso del tiempo de *stand-by* del *split* j^{th}
$\pi_i(t)$	Probabilidad del estado (i,t)
	Hay i tramas en el nodo:
	(1) t impar, usando el *split* $j:\ j=\frac{t+1}{2},\ (i,j)$
	(2) t par, *stand-by* tras *split* $j:\ j=\frac{t}{2},\ (i,\widetilde{j})$
π_i	Vector columna: $[\pi_i(1)\dots\pi_i(t)\dots\pi_i(2s)]$
$\mathcal{Q}$	matriz infinitesimal del proceso de QBD
F	matriz de re-envío
B	Matriz de transición hacia atrás
L, L_0	matrices de transición de estado con mismo número de tramas
	Red *fronthaul*
λ	Tasa de llegada el nodo (*switch*/enlace)
μ	Tasa de servicio del nodo (*switch*/enlace)
ρ	Ocupación del nodo (*switch*/enlace)
Λ	Vector de tasas de entrada en los nodos
Γ	Vector de tráfico externo
	$\gamma_i \neq 0$ solo para nodos CU
$\mathcal{R}$	Matriz de encaminamiento de la red *fronthaul*

A. Modelo de los nodos CU y DU

Como se ha mencionado anteriormente, el modelo de los nodos CU y DU es una extensión del presentado en [3]. Se considera una comunicación *downlink*, aunque se podría aplicar también al *uplink*. Se asume que las tramas llegan al CU siguiendo un proceso de *Poisson* de tasa $\lambda\ ms^{-1}$ y que se admiten s *splits*, cada uno de los cuales se caracteriza por un tiempo de servicio con distribución exponencial y valor medio $\mu_k^{-1}\ ms$ para cada *split* k^{th}. De acuerdo a la política adoptada se asume que el tiempo de permanencia en cada nivel de *split* k también está distribuido exponencialmente, con media $\gamma_k^{-1}\ ms$. Al cambiar de *split* se permanece en situación de *standby* durante el tiempo necesario para proceder a la reconfiguración, que también se asume exponencial, con media $\xi_k^{-1}\ ms$, para cada *split* k^{th}. Al abandonar un *split* k, y siempre de acuerdo con la política utilizada, el sistema usa el *split* l con probabilidad α_{kl}, y se define $\alpha_{kk}=0$, para asegurar que no se transita al mismo *split*.

Las principales mejoras con respecto al trabajo presentado en [3] son:

- Se consideran diferentes tiempos de permanencia para cada nivel de *split*.
- El tiempo en el estado *standby* es diferente para cada *split*.
- Se asegura que el *split* de destino es diferente al de origen en cada cambio.

Estas modificaciones permiten un modelado más realista y un mayor nivel de configuración, que da lugar a la cadena de *Markov* tri-dimensional que se muestra en la Figura 1.

Se definen dos tipos de estados con las duplas (i,j) y $(i,\widetilde{j})$ respectivamente. La primera dupla indica el estado de operación normal, donde i se corresponde con el número de tramas en el nodo, y j indica el índice asociado al *split* funcional. La segunda dupla representa el estado de *standby* al abandonar el *split* j. De este modo, la cadena de *Markov* resultante tiene s planos horizontales, cada uno representando un nivel de *split*. Si el nodo se encuentra activo en el *split* j, ante la llegada de una trama (tasa λ) o al terminar el procesado (tasa μ_j) se produce una transición a derecha o izquierda, respectivamente. Además, se puede transitar al estado de *stanby*, $(i,\widetilde{j})$, con tasa γ, de modo que las tramas pueden seguir llegando, pero no se procesan hasta abandonar ese estado, tal como se ve en la Figura 1.

Tras salir de *stanby* el nodo pasa a otro nivel de *split* con tasa ξ_j. En concreto el nuevo *split* k se selecciona con probabilidad α_{jk} con $k\in\{1,\dots,s\},\ k\neq j$, de modo que la tasa de transición desde $(i,\widetilde{j})$ a (i,k) es $\alpha_{jk}\cdot\xi_j$. Aunque durante la estancia en *standby* no se pueden procesar tramas, se asume que es posible almacenarlas hasta volver a estar en situación de atenderlas. El modelo presentado da lugar a un proceso de quasi nacimiento y muerte (quasi-birth-death, QBD), en el que cada nivel se corresponde con todos los estados con el mismo número de tramas: (i,j) y

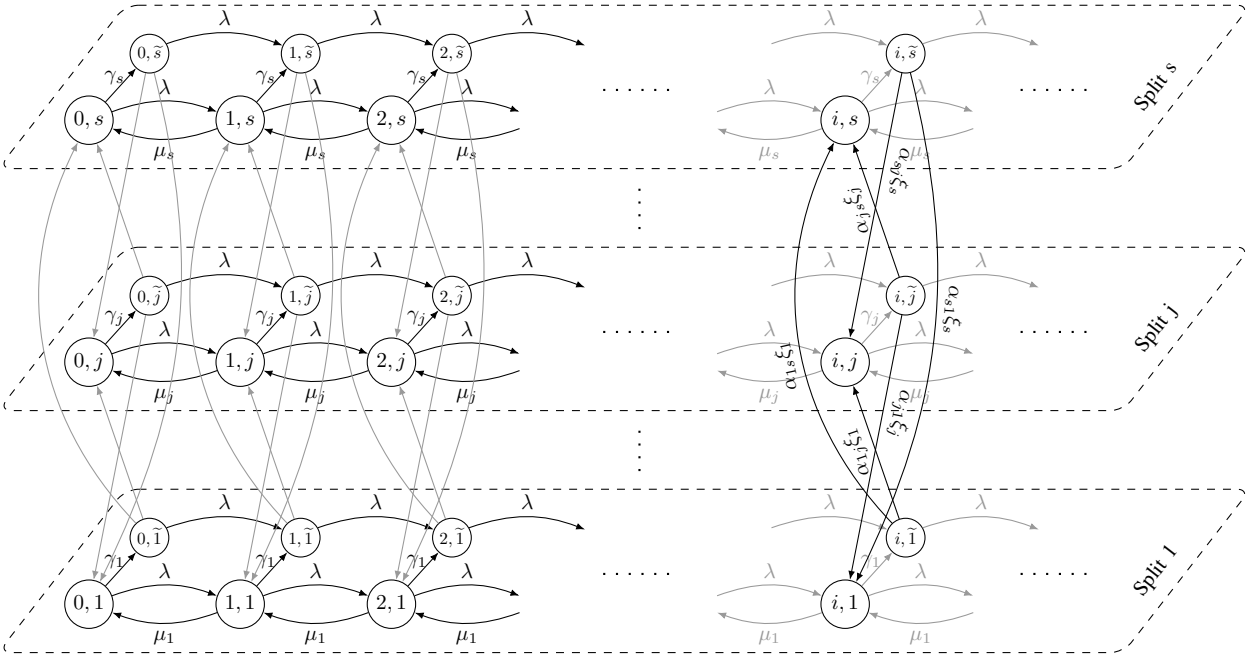


Fig. 1: Cadena de *Markov* para los nodos CU y DU

$(i, \widetilde{j})$, para $j, \widetilde{j} \in \{1, \ldots, s\}$. Se propone aplicar el método Matrix Geometric para definir el retardo de procesado de cada trama, tal como se muestran los trabajos de Neuts y Hajek [19], [20]. La matriz infinitesimal que caracteriza el QBD se define como:

$$\mathcal{Q} = \begin{bmatrix} L_0 & F & 0 & 0 & \cdots \\ B & L & F & 0 & \cdots \\ 0 & B & L & F & \cdots \\ \vdots & & \ddots & & \ddots \end{bmatrix} \tag{1}$$

donde $L_0, B, L, F \in \mathbb{R}^{2s \times 2s}$. Las matrices B, F se definen en la ecuación (2), L en la ecuación (3) y $L_0 = L + B$.

$$F = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{bmatrix},$$

$$B = \begin{bmatrix} \mu_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \mu_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \mu_s & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{2}$$

Se define la distribución estacionaria del proceso QBD como $\Pi = [\pi_0, \pi_1, \pi_2, \ldots]$, donde π_i es un vector columna de longitud $2s$, y $\pi_i(t),\ t \in \{1, \ldots, 2s\}$ es la probabilidad de tener i tramas en el nodo cuando: (1) t es impar, el nodo se encuentra en el *split* j, y $j = \frac{t+1}{2}$, (2) t es par, el nodo está en *standby* tras pasar por el *split* j, $j = \frac{t}{2}$. Si el nodo está trabajando en régimen de estabilidad, la distribución estacionario existe, y hay una matriz constante R que cumple la siguiente relación [19, Theorem 3.1.1]

$$R^2 \cdot B + R \cdot L + F = 0, \tag{4}$$

donde $R \in \mathbb{R}^{2s \times 2s}$. Aunque no hay una solución cerrada para la ecuación cuadrática (4), se puede utilizar un método iterativo para encontrar R. Además, se sabe que existe una única solución positiva, con la que se puede obtener π_0:

$$\begin{aligned} \pi_0^\intercal \left(L_0 + R \cdot B\right) &= \mathbf{0}^\intercal, \\ \pi_0^\intercal \left(I - R\right)^{-1} \mathbf{1} &= 1, \end{aligned} \tag{5}$$

donde $\mathbf{0}, \mathbf{1}$ son vectores de ceros y unos respectivamente, de longitud $2s$. Así, la distribución estacionaria $\mathbf{\Pi} = [\pi_0, \pi_1, \ldots]$ se obtiene como:

$$\pi_i^\intercal = \pi_0^\intercal \cdot R^i. \tag{6}$$

A partir de la distribución estacionaria, se pude obtener fácilmente el número medio de tramas en el nodo $\overline{N_{cu/du}}$ como:

$$\overline{N_{cu/du}} = \left\| \frac{\pi_1}{(I-R)^2} \right\|_1 = \left\| \frac{\pi_0^\intercal \cdot R}{(I-R)^2} \right\|_1 \tag{7}$$

donde $\| \cdot \|_1$ es la norma-1. Finalmente, usando la ley de *Little* se puede encontrar el retardo medio por trama $\tau_{\text{cu/du}}$, que tiene en cuenta tanto el tiempo de espera como de procesado:

$$\tau_{\text{cu/du}} = \frac{\overline{N_{\text{cu/du}}}}{\lambda} \tag{8}$$

Como se ha mencionado, la distribución estacionaria existe si la tasa de servicio media en el nodo es superior a la tasa de entrada. Por lo tanto, se puede establecer la

$$L = \begin{bmatrix} -(\lambda+\mu_1+\gamma_1) & \gamma_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda+\xi_1) & \alpha_{12}\cdot\xi_1 & 0 & \alpha_{13}\cdot\xi_1 & \cdots & \alpha_{1s}\cdot\xi_1 & 0 \\ 0 & 0 & -(\lambda+\mu_2+\gamma_2) & \gamma_2 & 0 & \cdots & 0 & 0 \\ \alpha_{21}\cdot\xi_2 & 0 & 0 & -(\lambda+\xi_2) & \alpha_{23}\cdot\xi_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & -(\lambda+\mu_s+\gamma_s) & \gamma_s \\ \alpha_{s1}\cdot\xi_s & 0 & \alpha_{s2}\cdot\xi_s & 0 & \alpha_{s3}\cdot\xi_s & \cdots & 0 & -(\lambda+\xi_s) \end{bmatrix} \tag{3}$$

tasa máxima $\lambda_{\max}$ que garantice la estabilidad del sistema como:

$$\lambda_{\max} = \sum_{i=1}^{s} \theta_i \cdot \mu_i \tag{9}$$

donde θ_i es la probabilidad de que el nodo se encuentre en un nivel de *split*, la cual se puede calcular resolviendo el siguiente sistema:

$$\Theta^{\intercal} \cdot A = \mathbf{0}^{\intercal} \quad ; \quad \Theta^{\intercal} \cdot \mathbf{1} = 1 \tag{10}$$

donde Θ es un vector columna de longitud $2s$ con las probabilidades de los *splits* y estado de *standby*, $A = L + B + F$, y $\mathbf{0}$ y $\mathbf{1}$ son vectores columna de ceros y unos respectivamente de longitud $2s$.

B. Retardo extremo a extremo en el fronthaul

Como se ha mencionado, se asumen que los nodos CU y DU están conectados por una red de conmutación de paquetes formada por *switches* y enlaces que podrían utilizar tecnologías diferentes. Estos elementos (enlaces y *switches*) se modelan como sistemas M/M/1, lo que permite aplicar teoría de redes abiertas de *Jackson*. Se modela la topología de red como un grafo dirigido $\mathcal{G} = (\mathbb{V}, \mathbb{E})$, donde $\mathbb{V}$ y $\mathbb{E}$ son el conjunto de nodos y enlaces, respectivamente. Si se asume que existen c CUs, d DUs, n *switches* y l enlaces, se puede definir $V \triangleq |\mathbb{V}| = c + d + n + l$. Se define la matriz de encaminamiento $\mathcal{R}$, de tamaño $V \times V$, que indica cómo las tramas recorren la red entre los CUs y sus DUs correspondientes. En la Figura 2 se muestra, a modo de ejemplo, una conexión entre la CU_x y DU_x a través de in *switch* S_x y los enlaces correspondientes. Como se puede ver, el modelo basado en el proceso QBD se utiliza en los nodos CU y DU, mientras que el *switch* y los enlaces se modelan como sistemas M/M/1.

Si se asume que se respetan las condiciones de los teoremas de Burke y Jackson [21], [22], se puede establecer el retardo extremo a extremo como la suma de los retardos asociados a cada nodo en la ruta. Estas condiciones implican que el proceso de tráfico a la salida de cada nodo sea estadísticamente idéntico al de entrada. En el caso de los nodos M/M/1 el retardo se puede calcular como $\tau_{\text{mm1}} = \frac{1}{\mu-\lambda}$, donde μ y λ son las tasas de servicio y de tráfico de entrada al nodo. En este caso, se garantiza la estabilidad si $\mu > \lambda$. Se asume que únicamente los CUs reciben tráfico, y que la matriz de encaminamiento $\mathcal{R}$ indica la ruta hasta el DU correspondiente. Además, los *switches* y enlaces pueden ser compartidos por varios flujos de tráfico. Con ello, se define Λ como el vector de tasas de entrada λ_v de cada nodo $v \in \mathbb{V}$, que se puede calcular como [23], [22]:

$$\Lambda = \Phi \cdot (\mathcal{I} - \mathcal{R})^{-1} \tag{11}$$

donde Φ es otro vector fila que contiene el tráfico externo en la red, de modo que $\phi_v = 0$ para los *switches*, enlaces y DUs, y $\phi_v \neq 0$ para los CUs. Por lo tanto, usando la matriz de encaminamiento $\mathcal{R}$ y las tasas de entrada en los CUs se puede calcular la la tasa de entrada y la ocupación en cada nodo y, a partir de ello, el retardo correspondiente. Finalmente, el retardo extremo a extremo para cada flujo $f \in \mathbb{F}$, siendo $\mathbb{F}$ el conjunto de flujos de entrada, se obtiene como:

$$\overline{\tau_f} = \sum_{v \in \mathcal{P}(f)} \tau_v \quad ; \quad \mathcal{P}(f) : \mathbb{F} \longrightarrow \mathbb{V} \tag{12}$$

donde $\mathcal{P}(f)$ es una función que devuelve los nodos que atraviesa el flujo f.

Además, es posible establecer el retardo promedio en la red (sin necesidad de calcular los retardos individuales por flujo), aplicando la ley de *Little*:

$$\overline{\tau} = \frac{\sum_{v \in \mathbb{V}} n_v}{\lambda_0} \tag{13}$$

donde λ_0 es la tasa total de tráfico externo en la red: $\lambda_0 = \sum_{v \in \mathbb{V}} \phi_v$. Por otro lado n_v es el número medio de tramas en el nodo v, definido (para *switches* y enlaces) por:

$$n_v = \frac{\rho}{1-\rho} \tag{14}$$

En la ecuación (14) ρ representa la ocupación del nodo, calculada como $\rho = \frac{\lambda}{\mu}$. Como se discutirá a continuación, el proceso de salida de las CUs y DUs no es estrictamente de *Poisson* y esto puede dificultar el uso de la teoría de redes abiertas de *Jackson*. Como se verá, bajo situaciones razonables (tiempos de *standby* pequeños), los resultados siguen siendo válidos y cercanos al rendimiento real.

IV. Validación del modelo

En esta sección se validará el modelo descrito anteriormente, comparando los resultados teóricos con los obtenidos mediante simulación. Para ello se hará uso de un simulador por eventos implementado en `C++`, que ha sido implementado *ad-hoc*. La razón de utilizar una nueva herramienta en lugar de soluciones existentes (p.e. `ns-3`) es que el objetivo es la validación del modelo, por lo que se ha buscado tener un mayor control sobre el

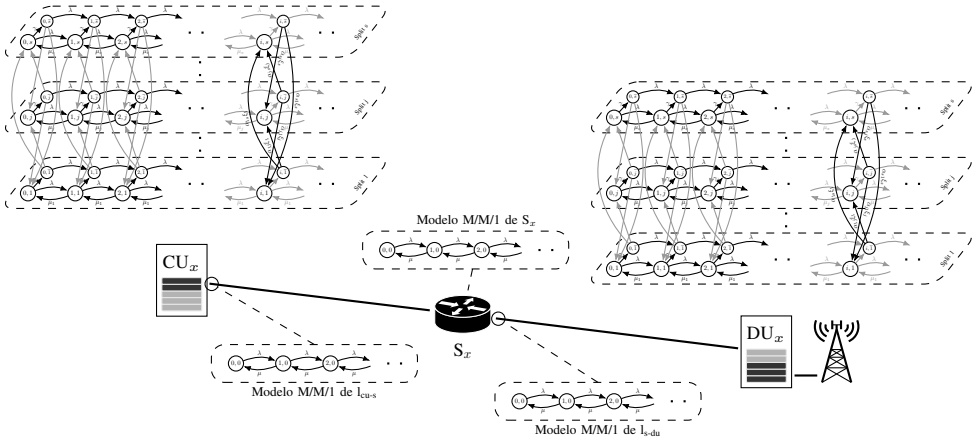


Fig. 2: Modelo extremo a extremo basado en cadenas de *Markov*

comportamiento simulado y evitar la complejidad añadida por soluciones más completas, tales como lógica de protocolos, que tendrían algún impacto en los resultados. A modo de resumen, el simulador implementa los dos tipos de nodo utilizados (M/M/1 y QBD) y cuatro tipos de eventos. Todos los nodos gestionan dos clases de eventos: (1) llegada de una trama y (2) finalización de procesado de trama. Además, en los nodos QBD hay otros dos tipos de eventos: (3) cambio de *split* y (4) finalización de *stanby*. Se pueden configurar varios flujos de entrada, y la matriz de encaminamiento indica las rutas que las tramas pertenecientes a estos flujos siguen. En la Tabla II se muestran los parámetros de configuración utilizados en todos los escenarios. Se han considerado 4 niveles de *split* ($s = 4$), con tasas de servicio $\mu_{1,2,3,4} = \{1, 1.5, 2, 4\}$ ms^{-1}. Estos valores se han seleccionado para ilustrar el potencial del modelo, y reflejan diferentes capacidades de procesado de tramas de los niveles de *split*.

Por otro lado, el tiempo medio de permanencia en cada *split* es $\gamma_{1,2,3,4} = \{\frac{1}{100}, \frac{2}{100}, \frac{3}{100}, \frac{4}{100}\}$ ms^{-1}. Como se puede ver, las tasas para las DUs son "complementarias", ya que el procesado total ha de repartirse entre la CU y la DU. Además, se asume que $\xi_j = \xi \ \forall j$ de modo que el tiempo de *standby* es el mismo para todos los *splits*. La matriz A establece las probabilidades de selección del siguiente nivel de *split*, siendo la probabilidad de transitar al mismo estado $\alpha_{i,i} = 0$, y asegurando que una vez iniciado el cambio de *split* este finalice, $\sum_{j=1^s} \alpha_{i,j} = 1$. Como se puede observar en la Tabla II la matriz A en los DUs también es la "complementaria" de la correspondiente a los CUs, para reflejar los cambios de *split* correspondientes.

En cuanto a los nodos M/M/1, utilizados para modelar los *switches* y enlaces, se han definido varias tasas de servicio, para reflejar diferentes situaciones y tecnologías. Inicialmente la tasa de servicio de los *switches* será $\mu_n = 5\ ms^{-1}$ y se reducirá a 3 ms^{-1} en el último escenario. Asimismo, las tasas de los enlaces representan dos tecnologías: fibra óptica con una tasa de $\mu_{\text{of}} = 8\ ms^{-1}$ y ondas milimétricas, cuya tasa μ_{mmw} se variará (1, 2, 4, 6 ms^{-1}) para analizar su impacto.

Tabla II: Configuración del escenario

Nodos CU y DU	
Tasas de servicio	$\mu = \{1, 1.5, 2, 4\}$ (ms^{-1})
Tasas de cambio de *split*	$\gamma_{\text{cu}} = \{\frac{1}{100}, \frac{2}{100}, \frac{3}{100}, \frac{4}{100}\}$ (ms^{-1})
	$\gamma_{\text{du}} = \{\frac{4}{100}, \frac{3}{100}, \frac{2}{100}, \frac{1}{100}\}$ (ms^{-1})
Duración de *standby*	$\xi^{-1} = 1, 5, 10, 20, 50$ (ms)
Probabilidades de transición entre *splits*	$A_{\text{cu}} = \begin{pmatrix} 0 & 0.6 & 0.2 & 0.2 \\ 0.1 & 0 & 0.3 & 0.6 \\ 0.3 & 0.3 & 0 & 0.4 \\ 0.2 & 0.3 & 0.5 & 0 \end{pmatrix}$
	$A_{\text{du}} = \begin{pmatrix} 0 & 0.5 & 0.3 & 0.2 \\ 0.4 & 0 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0 & 0.1 \\ 0.2 & 0.2 & 0.6 & 0 \end{pmatrix}$
Red *fronthaul*	
Tasas de servicio de los *switches*	$\mu_n = 5, 3$ (ms^{-1})
Tasa de servicio de enlaces de fibra óptica	$\mu_{\text{of}} = 8$ (ms^{-1})
Tasa de servicio de enlaces mmWave	$\mu_{\text{mmw}} = 1, 2, 4$ (ms^{-1})

A. *Nodos CU/DU*

En el primer escenario a evaluar se validará el comportamiento de los nodos CU y DU. Se usará la configuración indicada en la Tabla II y se estudiará el tiempo de permanencia en el nodo al incrementar la tasa de entrada para los diferentes valores de tiempo de *standby*. En la Figura 3 se muestran los resultados teóricos con línea continua, y los obtenidos con el simulador con marcadores. Los valores de simulación se han obtenido a partir de 100 simulaciones independientes, en cada una de las cuales se han generado 10^6 tramas, para asegurar resultados estadísticamente fiables. En primer lugar, se puede observar que con el modelo teórico se obtienen valores casi idénticos a los simulados, lo que permite validar el modelo de los nodos CU/DU, así como la correcta implementación del simulador. Por otro lado, los resultados también indican que el tiempo de *standby* tiene un gran impacto, ya que el tiempo medio de permanencia crece de forma acusada al aumentar el valor de ξ^{-1}. Merece la pena indicar que en sistemas reales, es esperable que el tiempo de *stanby* necesario para la reconfiguración de las estaciones base sea varios órdenes de magnitud menor que el de permanencia en cada uno de los *split*,

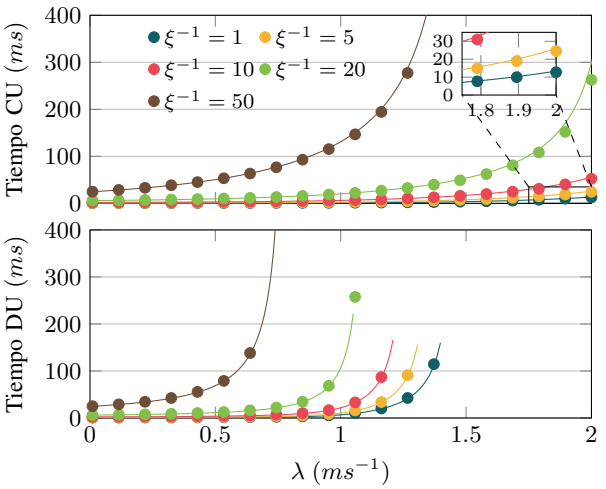


Fig. 3: Tiempo de permanencia en los nodos CU/DU incrementando la tasa de entrada λ y con diferentes tiempos de *standby*

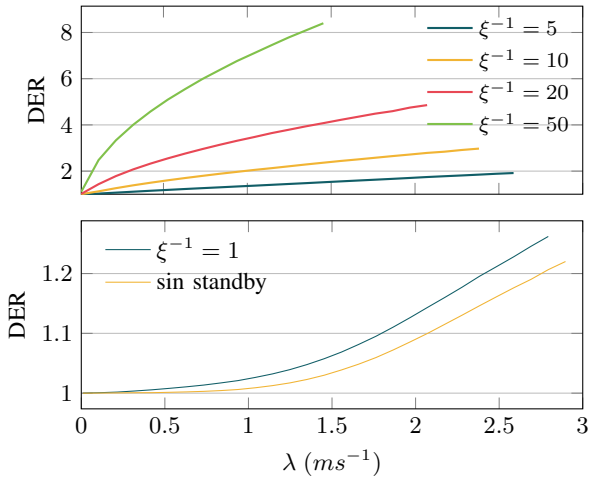


Fig. 4: Desviación estándar relativa (DER) del tiempo entre salidas en el CU para diferentes valores de la tasa de entrada y tiempos de *standby*

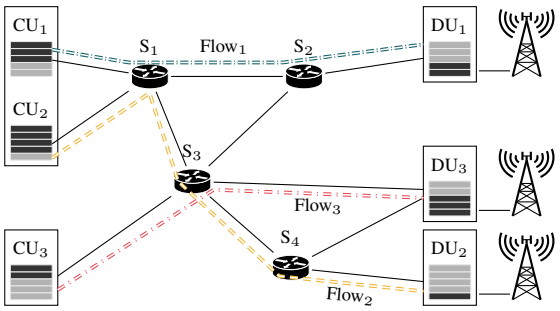


Fig. 5: Red *fronthaul* para validar el modelo extremo a extremo

como se ha visto en [5].

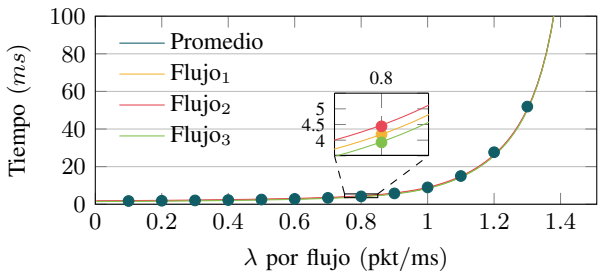
Como se ha mencionado previamente, para utilizar la teoría de redes abiertas de *Jackson* en la caracterización del retardo extremo a extremo se requiere que se cumpla el teorema de Burke, de modo que el proceso de tráfico a la salida sea estadísticamente idéntico al de la entrada [21], [23]. Por ello, se necesita asegurar que el tráfico de salida en los CU sea un proceso de *Poisson*, o de otro modo, que el tiempo entre salidas consecutivas sigue una distribución exponencial. Incluso si el tráfico de entrada sea tal que se asegure la estabilidad del sistema, definida por la ecuación (9), podría haber circunstancias en las que el teorema de Burke no se cumpliera. Por ello, se debe asegurar que: (i) el tráfico de entrada sea menor que la tasa de servicio del *split* más lento y (ii) que el tiempo de *standby* pueda considerarse despreciable en comparación con los tiempos de permanencia en los *split*.

A fin de analizar si estas dos condiciones han de respetarse de manera estricta, se ha usado el simulador para estudiar el tiempo entre salidas en el CU. En la Figura 4 se muestra la desviación estándar relativa (DER) de estos tiempos, que se define como la relación entre su desviación estándar y su media. Si la salida del nodo CU fuera un proceso de *Poisson*, la DER debe tomar valor 1. Se puede observar que la DER es notablemente mayor que 1 cuando el tiempo de *standby* es alto, por lo que en esas circunstancias el proceso de salida del tráfico no podría ser considerado de *Poisson*, incluso cuando la tasa de entrada está por debajo de su posible valor máximo, aquel que asegura estabilidad. Por otro lado, cuando el valor del tiempo de *standby* es menor, la DER está muy próxima a la unidad. Dado que esta situación es la más verosímil, se puede considerar que en condiciones realistas el tráfico a la salida del CU se corresponderá con un proceso de *Poisson*, y que por lo tanto el modelo presentado será válido.

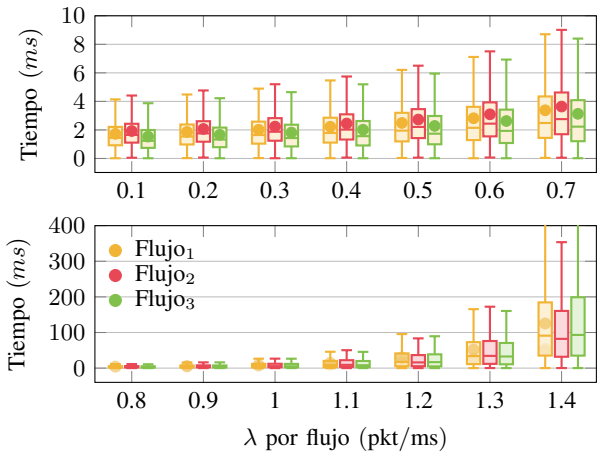
B. Retardo extremo a extremo

A continuación se analizará el retardo extremo a extremo esperado sobre el escenario representado en la Figura 5 que incluye tres pares CU/DU y cuatro *switches*. Sobre este escenario se establece un flujo entre cada par CU/DU, el cual sigue la ruta mostrada en la figura: (i) $CU_1 \rightarrow S_1 \rightarrow S_2 \rightarrow DU_1$; (ii) $CU_2 \rightarrow S_1 \rightarrow S_3 \rightarrow S_4 \rightarrow DU_2$; (iii) $CU_3 \rightarrow S_3 \rightarrow DU_3$.

En primer lugar se asume que todos los enlaces tienen alta capacidad (fibra óptica), por lo que únicamente se incluye en la evaluación del retardo el efecto de los nodos CU/DU y de los *switches*. En la Figura 6a se muestran los retardos medios de cada flujo y el promedio global, al incrementar el valor de la tasa de entrada. Los resultados teóricos se han obtenido con las ecuaciones (13) y (14), y se han comparado con los obtenidos mediante simulación. Nuevamente, para cada configuración se han realizado 100 simulaciones independientes generando en cada una de ellas 10^6 tramas. Como se puede observar, nuevamente los resultados teóricos son casi idénticos a los simulados, aumentando el retardo en ambos casos al incrementar la tasa de entrada. También se puede observar que el modelo teórico proporciona resultado prácticamente idénticos, en su valor medio, a los simulados, incluso cuando el tráfico de entrada es superior a la tasa de servicio del *split* más lento (1 pkt/ms), lo que implica que no se



(a) Retardo medio



(b) Distribución del retardo

Fig. 6: Retardo extremo a extremo al incrementar λ por flujo. La figura superior muestra el retardo medio y la inferior representa la variabilidad de los resultados obtenidos.

cumplen estrictamente los requisitos para aplicar la teoría de *Jackson*.

Usando el simulador se puede extender el análisis para conocer, no solo los valores medios, sino la distribución del retardo, que puede tener un impacto notable en el rendimiento de los servicios. En la Figura 6b se usan diagramas de caja (*boxplots*) para representar la variabilidad del retardo por flujo para varios valores de tasa de entrada (λ). Cada diagrama indica la mediana (percentil del 50%) con una línea horizontal, así como los percentiles del 25 y 75%, que corresponden a los límites de la cada. Por otro lado, las líneas superior e inferior indican los percentiles del 5 y 95%. También se indica en cada caja el valor medio mediante un marcador. Como se puede observar el retardo crece al aumentar la tasa de entrada, tal como se vio anteriormente. Estos resultados muestran que para tasas de entrada bajas el retardo máximo se encuentra por debajo de 10ms, el cual aumenta bruscamente cuando la tasa de entrada supera la tasa de servicio del *split* más lento (1 pkt/ms).

En el siguiente escenario se fija la tasa de los flujos 1 y 3 (f_1 y f_3) a 0.8 pkt/ms, y se va incrementando la correspondiente al flujo 2 (f_2). Como se puede apreciar en la Figura 5, f_2 atraviesa los *switches* S_1, que también es usado por f_1, y S_3, que se comparte con f_3. En la Figura 7 se muestra el retardo medio extremo a extremo. Al igual

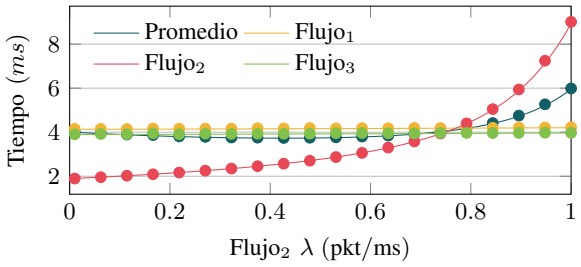


Fig. 7: Retardo extremo a extremo incrementando λ_{f_2}

que en los resultados anteriores los valores teóricos se indican con línea continua, mientras que las resultados obtenidos en la simulación se representan con marcadores, que indican el valor medio tras 100 experimentos independientes. En este caso, dado que las tasas de los flujos son diferentes, cada simulación genera tramas hasta asegurar que el flujo con menor tasa envía 10^6 tramas, y que el resto no ha dejado de generar tráfico, para que las condiciones de la red no cambien a lo largo de cada experimento. Además de comprobarse nuevamente que los resultados de la simulación y teóricos son casi idénticos, se puede observar que el incremento de la tasa λ_{f_2} prácticamente no influye en los otros flujos con los que comparte *switch*, ya que los *switches* en este escenario tienen poca ocupación en relación a su capacidad. Por otro lado, los resultados también muestran el incremento del retardo de f_2 aumentar su tasa, por lo que se deduce que, con esta configuración, el retardo es debido principalmente al procesado en los nodos CU/DU.

C. Impacto de estrategia de encaminamiento y enlaces heterogéneos

Se analiza a continuación el impacto sobre el retardo al modificar la tecnología de los enlaces que conforman la red que se está analizado (Figura 5), así como al adaptar la configuración de encaminamiento. Se asume que todos los enlaces tienen capacidad alta ($\mu_{\text{fo}} = 8\ ms^{-1}$), excepto el que conecta los *switches* S_1 y S_2, que emula un enlace mmWave. Bajo estas condiciones se ha variado la política de encaminamiento de S_1, de modo que con probabilidad φ se usa el camino corto (atravesando el enlace entre S_1 y S_2) y con probabilidad $1-\varphi$ el tráfico se reenvía por la siguiente ruta: $\text{CU}_1 \rightarrow S_1 \rightarrow S_3 \rightarrow S_2 \rightarrow \text{DU}_1$. La Figura 8 muestra que el retardo medio global (considerando todos lo flujos) varía a medida que se modifica el valor de φ. Las tasas para todos los flujos son 0.8 pkt/ms, y los resultados se representan como en las figuras anteriores. Los valores que se obtienen a través del simulador también se han obtenido de 100 simulaciones independientes, en las que se han generado 10^6 tramas por flujo. Los resultados muestran que la estrategia de encaminamiento, como era de esperar, tiene un impacto evidente en el rendimiento. En este sentido, se puede ver que hay un punto de operación óptimo (respecto a φ) donde el retardo presente el valor más bajo. En concreto, con la configuración descrita, cuando la tasa del enlace entre los *switches* S_1 y S_2 es 1 pkt/ms, el valor que optimiza el rendimiento es $\varphi \approx 0.6$.

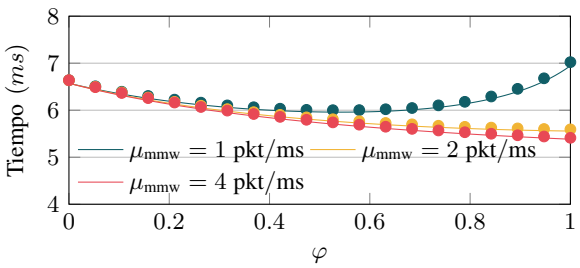


Fig. 8: Retardo extremo a extremo con enlaces heterogéneos y diferentes políticas de encaminamiento

V. Conclusiones

En este artículo se ha presentado un modelo basado en teoría de colas para analizar el rendimiento de la red *fronthaul* en redes de acceso con selección dinámica de *functional split*. El modelo considera diferentes tasas de servicio para cada uno de los niveles de centralización, así como diferentes tiempos de *standby*, que se contempla para las tareas de reconfiguración de los nodos CU y DU al modificar su split. Se plantea el uso de un proceso QBD, cuyo comportamiento se ha obtenido usando el método de la matriz geométrica. También se ha analizado bajo qué circunstancias el modelo de los nodos CU/DU se puede utilizar junto con teoría de redes de *Jackson* para evaluar el retardo extremo a extremo en la red *fronthaul*. Se ha visto que, para regímenes de operación realistas, los resultados obtenidos por el modelo teórico son casi idénticos a los proporcionados mediante simulación.

Posteriormente se ha estudiado el retardo extremo a extremo de la red *fronthaul*, observando nuevamente que los resultados proporcionados por el modelo y los obtenidos mediante simulación son prácticamente idénticos. Finalmente, se ha modificado la configuración del escenario de evaluación para mostrar el potencial del modelo ante diferentes circunstancias. En concreto, se ha analizado el rendimiento de la red al aplicar diferentes políticas de encaminamiento sobre enlaces heterogéneos, poniendo de manifiesto que el modelo puede ser utilizado para obtener puntos óptimos de operación.

Se han identificado dos líneas de trabajo que se abordarán en el futuro. Por un lado, utilizando el simulador se va a analizar el impacto que tiene limitar el tamaño de los *buffer* en los diferentes nodos, así como el efecto de cambiar los patrones de tráfico. Por otro lado, se pretende utilizar el modelo para evaluar diferentes políticas de *split*, y esquemas de gestión de los *buffer*.

Agradecimientos

Los autores agradecen la financiación de Gobierno de España (Ministerio de Economía y Competitividad, Fondo Europeo de Desarrollo Regional, MINECO-FEDER) por medio del proyecto *FIERCE: Future Internet Enabled Resilient smart CitiEs* (RTI2018-093475-AI00).

References

[1] C. I, Y. Yuan, J. Huang, S. Ma, C. Cui, and R. Duan, "Rethink fronthaul for soft ran," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 82–88, Sep. 2015.

[2] G. O. Pérez, J. A. Hernández, and D. Larrabeiti, "Fronthaul network modeling and dimensioning meeting ultra-low latency requirements for 5g," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 6, pp. 573–581, June 2018.

[3] L. Diez, C. Hervella, and R. Agüero, "Understanding the performance of flexible functional split in 5g vran controllers: A markov chain-based model," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 456–468, 2021.

[4] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[5] A. M. Alba, J. H. G. Velásquez, and W. Kellerer, "An adaptive functional split in 5g networks," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 410–416.

[6] C.-Y. Chang, N. Nikaein, R. Knopp, T. Spyropoulos, and S. S. Kumar, "Flexcran: A flexible functional split framework over ethernet fronthaul in cloud-ran," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–7.

[7] D. Harutyunyan and R. Riggio, "Flex5g: Flexible functional split in 5g networks," *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 961–975, 2018.

[8] ——, "Flexible functional split in 5g networks," in *2017 13th International Conference on Network and Service Management (CNSM)*, 2017, pp. 1–9.

[9] V. Q. Rodriguez, F. Guillemin, A. Ferrieux, and L. Thomas, "Cloud-ran functional split for an efficient fronthaul network," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, 2020, pp. 245–250.

[10] Y.-T. Huang, C.-H. Fang, L.-H. Shen, and K.-T. Feng, "Optimal functional split for processing sharing based comp for mixed embb and urllc traffic," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[11] A. Martinez Alba and W. Kellerer, "A dynamic functional split in 5g radio access networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[12] A. Alabbasi, M. Berg, and C. Cavdar, "Delay constrained hybrid cran: A functional split optimization framework," in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–7.

[13] T. Ismail and H. H. M. Mahmoud, "Optimum functional splits for optimizing energy consumption in v-ran," *IEEE Access*, vol. 8, pp. 194 333–194 341, 2020.

[14] L. Wang and S. Zhou, "Flexible functional split and power control for energy harvesting cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1535–1548, 2020.

[15] H. Gupta, M. Sharma, A. Franklin A., and B. R. Tamma, "Apt-ran: A flexible split-based 5g ran to minimize energy consumption and handovers," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 473–487, 2020.

[16] S. Zhou, X. Liu, F. Effenberger, and J. Chao, "Mobile-pon: A high-efficiency low-latency mobile fronthaul based on functional split and tdm-pon with a unified scheduler," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, 2017, pp. 1–3.

[17] A. Marotta, D. Cassioli, K. Kondepu, C. Antonelli, and L. Valcarenghi, "Efficient management of flexible functional split through software defined 5g converged access," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[18] M. P. Amaral, J. Gomes, H. R. O. Rocha, J. A. L. Silva, and M. E. V. Segatto, "Processing resource allocation in 5g fronthaul," in *2019 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, 2019, pp. 1–3.

[19] M. Neuts, "Markov Chains with Applications in Queueing Theory, Which Have a Matrix-Geometric Invariant Probability Vector," *Advances in Applied Probability*, vol. 10, no. 1, pp. 185–212, 1978.

[20] B. Hajek, "Birth-and-death processes on the integers with phases and general boundaries," *Journal of Applied Probability*, vol. 19, no. 3, p. 488–499, 1982.

[21] P. J. Burke, "The Output of a Queuing System," *Operations Research*, vol. 4, no. 6, 1956.

[22] J. R. Jackson, "Jobshop-like Queueing Systems," *Management Science*, vol. 10, no. 1, 1963.

[23] L. Kleinrock, *Queueing Systems. Volume 1: Theory*. Wiley-Interscience, 1975.

XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Rendimiento de Redes 4G/5G usando una estación base real

Félix Delgado-Ferro, Jorge Navarro-Ortiz, Lorena Chinchilla-Romero, Pablo Muñoz-Luengo
Departamento de Teoría de la Señal, Telemática y Comunicaciones,
Universidad de Granada
C/ Periodista Daniel Saucedo Aranda, s/n. ETSI Informática y de Telecomunicación.
felixdelgado@correo.ugr.es, jorgenavarro@ugr.es, lorenachinchilla@ugr.es, pabloml@ugr.es

Este artículo describe el desarrollo del proyecto sobre el despliegue de red móvil 5G y análisis de características de la misma. Actualmente, se encuentra en desarrollo y trata del análisis del rendimiento y comparación en redes 4G y 5G empleando una estación base real. En este trabajo se incluye la información sobre el estudio previo de estas redes, las alternativas de implementación, el despliegue realizado empleando el software de Amarisoft y testeo de capacidades de las redes propuestas.

***Palabras Clave* – Redes móviles, 4G, 5G, estación base**

I. INTRODUCCIÓN

Actualmente, y desde 2017, nos encontramos en pleno auge de la conocida como quinta generación de comunicaciones móviles (5G) que fue impulsada por la estandarización (*Release 15*) de 3GPP [1] que recoge las técnicas y mecanismos para obtener los requisitos definidos por la ITU para IMT-2020. Estos requisitos pretenden ofrecer mejoras en relación a tres tipos de servicios:

- *Enhanced Mobile Broadband* (eMBB): pretende mejorar la capacidad de la red debido al incremento de la demanda producida por los usuarios. Es decir, se centra en mejorar las características ofrecidas por 4G para las comunicaciones entre usuarios.
- *Massive Machine Type Communications* (mMTC): pretende ofrecer comunicaciones eficientes, fiables y seguras para comunicaciones entre máquinas a gran escala.
- Ultra Reliable and Low Latency Communications (URLLC): ofrece comunicaciones con requisitos estrictos respecto a latencia y fiabilidad para abrir el abanico de aplicaciones posibles (e.g. V2x).

En relación a estos avances, el trabajo propuesto consiste en el despliegue de tres tipos de redes para telefonía móvil como 4G, 5G NSA y 5G SA. Una vez se hallan consolidado y probado la conectividad y funcionamiento, se realizarán una serie de pruebas en relación a la capacidad de transmisión que cada una de estas redes pueden ofrecernos empleando una estación base real.

II. ESTADO DEL ARTE

Hoy en día, tenemos múltiples soluciones para el despliegue y utilización de redes 4G y 5G en el ámbito de la investigación que nos permite configurar, modificar, simular y analizar redes móviles de forma realista. A continuación, se describen las plataformas más relevantes.

srsLTE [2] ofrece implementaciones de software modular y portable para tecnologías inalámbricas que se fundamentan en el software *Open-source srsRAN* que permite la implementación de los nodos de la red LTE (4G) por medio de los módulos srsLTE, srsUE y srsENB.

OpenAirInterface (OAI) [3] es una plataforma de código abierto que ofrece un entorno móvil flexible y de bajo coste para experimentación sobre redes 4G y 5G basándose en los estándares de 3GPP. Este compuesto por el software para desplegar la parte troncal de la red (CN) y la parte de acceso radio (RAN) por separado y con interconectividad entre ellas.

Free5gc [4] es una implementación open-source de las redes troncales 5G, tanto NSA (*non-stand alone*) como SA (*standalone*) siguiendo los estándares de 3GPP.

Amarisoft [5] es una empresa francesa que ha desarrollado soluciones asequibles y de alta calidad para el desarrollo de redes 4G/5G, proporcionando flexibilidad, disponibilidad y estabilidad mediante la implementación de módulos como: eNodeB, gNodeB, EPC, 5GC (*5G Core*) y UE Simulator.

En la figura 1, podemos ver los módulos desarrollados por Amarisoft para las implementaciones de las redes.

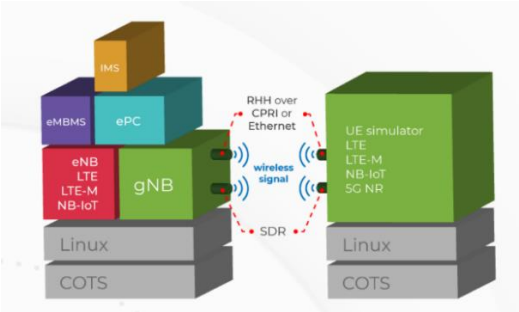


Fig. 1. Esquema de módulos de Amarisoft [5]

Al comprender la versatilidad y funcionamiento a grosso modo, se ha decidido emplear la plataforma de Amarisoft, en concreto la Serie AMARI NW [5], dado que esta es la solución más completa al ofrecer múltiples funcionalidades y configuraciones, siendo compatible con los estándares 3GPP.

Cabe destacar, que esta solución conlleva un coste económico superior al resto por tratarse de una plataforma comercial. Aún así, se ha optado por esta plataforma debido a disponer de estas estaciones base que son más fiables y teniendo como objetivo desarrollar un demostrador y un banco de pruebas (testbed) que fuese útil para futuros proyectos de investigación.

III. Fundamentos Teóricos

A. *Redes LTE (4G)*

Las redes móviles de cuarta generación (4G) han sido desplegadas empleando el estándar Long Term Evolution (LTE) del 3GPP. LTE es un estándar de comunicaciones móviles desarrollado en el Release 8 y mejorado en las siguientes revisiones hasta la Release 13.

La motivación de estas redes era simplificar las redes móviles y diferenciar las partes que las componen. Por tanto, el diseño de la arquitectura LTE pretendía optimizar el control del tráfico de datos y de los usuarios, optando por una arquitectura plana que minimizase el número de nodos implicados en la comunicación.

La arquitectura de red LTE, como apreciamos en la figura 2, se basa en tres componentes principalmente [6]:

- *Evolved Packet Core* (EPC): es un elemento fundamental dentro de la red e implementa el núcleo. La principal innovación consiste en emplear el protocolo IP para el transporte de todos los servicios. Además, otra diferencia relevante es que permite la separación de los planos de usuario (datos) y control (señalización). El EPC se compone principalmente de cuatro módulos:
  - *Mobility Management Entity* (MME): se encarga de la señalización de movilidad y seguridad de la red de acceso.
  - *Home Subscriber Server* (HSS): base de datos de la red y contiene la información del usuario y del control de movilidad.
  - *Serving Gateway* (S-GW): se encarga de transportar el tráfico de datos IP entre redes y gestionar la movilidad interna y externa.
  - *Packet Data Network Gateway* (P-GW): es el nodo de interconexión y se encarga de enrutar los paquetes de datos y gestionar los permisos y políticas sobre estos paquetes.
- *Evolved Universal Mobile Telecommunications System Terrestrial Radio Access Network* (E-UTRAN): es la red de acceso radio (RAN) y está compuesta por un conjunto de estaciones base (eNB) que pueden intercomunicarse entre sí para simplificar la señalización, el envío de mensajes y gestionar los recursos radio, la encriptación de paquetes, etcétera.
- *User Equipment* (UE): son los equipos de los usuarios y les permiten conectarse a la red.

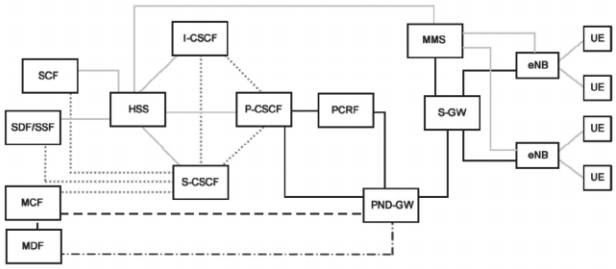


Fig. 2. Arquitectura de Red 4G [7]

B. *Redes NR (5G)*

Las especificaciones del núcleo de red 5G se encuentran en los estándares (*Release 15 y 16*) de 3GPP cuyo objetivo es dar cobertura a un abanico amplio de servicios y tecnologías como IoT, V2x, MTC, etcétera.

Las especificaciones principales recogen la integración de tecnologías como la virtualización de redes y network slicing y definen dos tipos de escenarios como se muestra en la figura 3:

- *Non Standalone* (NSA): integra dos generaciones de red de acceso (LTE y NR) conectadas a un mismo EPC. Ésta se compone de los eNB (nodos principales) y gNB (nodos secundarios) que comparten la gestión de tráfico de los usuarios, pero la gestión con el EPC la realiza únicamente el eNB.
- *Stand Alone* (SA): emplea la tecnología New Radio (NR) y se implementa en el Next Generation NodeB (gNB) y en el núcleo de la red (5GC). En este escenario, los nodos gNB que se encargan de las conexiones UE-5GC y de gestionar todo el tráfico.

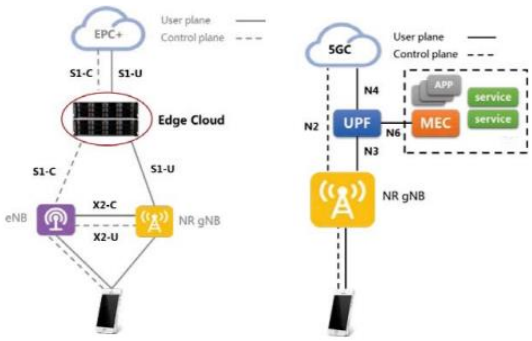


Fig. 3. Escenarios de despliegue de la red 5G (NSA y SA) [8]

La arquitectura de red 5G se divide en la parte de acceso radio (RAN) y el núcleo de la red (5GC).

La red de acceso radio se ve ampliada con el concepto de NFV y propone la segmentación de la red en Unidades Distribuidas (DU) empleando el interfaz normalizado *enhanced Common Public Radio* (eCPRI).

El núcleo de la red 5G permite procesar los datos de usuarios y gestionar la señalización de la red. El modelo 5GC consiste en una red integrada mediante interfaces punto a punto a una *Service Based Arquitecture* (SBA). El núcleo se establece mediante funciones como las que se muestran en la figura 4.

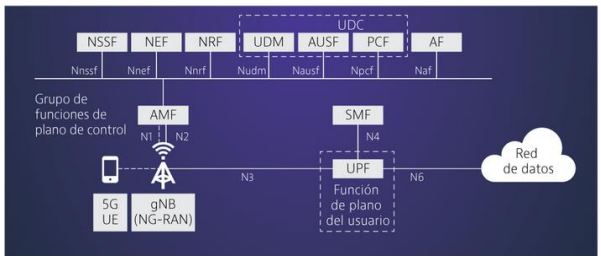


Fig. 4. Funciones de red elementales en una red 5G [9]

Las funciones fundamentales para el correcto funcionamiento del núcleo de la red 5G se exponen a continuación:

- *Access and Mobility Management Function* (AMF): se encarga del control general de la red e interacción radio RAN-UE (registro, autenticación y movilidad).
- *Session Management Function* (SMF): se encarga de la administración de las sesiones de los usuarios (establecimiento, modificación y liberación), así como de la asignación de IP a los dispositivos.
- *User Plane Function* (UPF): se encarga de procesar y reenviar el tráfico del plano de usuario (redirección y comunicación entre núcleo de red y red de datos). Además, realiza procedimientos de los datos reenviados como generar informes de tráfico, analizar el contenido de los paquetes de datos y ejecutar políticas.
- *Unified Data Management* (UDM): se encarga de gestionar la base de datos de los subscriptores móviles, las credenciales de autenticación y autorizar el acceso a la información disponible.
- *Authentication Server Function* (AUSF): es el proveedor del servicio de autenticación de los dispositivos.
- *NF Repository Function* (NRF): se encarga de registrar las funciones de la red realizadas por los servicios y que posibilitan las comunicaciones entre funciones.

IV. Despliegue de la Red 4G

La idea fundamental es partir de una red funcional 4G para posteriormente implementar la red 5G NSA y poder comparar características entre ellas. Posteriormente, se desplegará una red 5G SA de forma independiente.

Previo al despliegue, es decir, instalación del software Amarisoft, debemos preparar nuestro entorno de ejecución. En este proyecto se ha empleado un S.O de servidor Ubuntu 18 y se han preinstalado dos tarjetas PCIe SDR para la conectividad RAN. Posteriormente, se instala el software LTE 100 de Amarisoft.

En la configuración de los módulos, se tiene en cuenta la distribución de estos dependiendo de la parte de la red correspondiente. En el *Evolved Packet Core* (EPC) encontramos principalmente por dos módulos: MME e IMS, mientras que en la Radio Access Network (RAN) tenemos el módulo eNB.

Los módulos se pueden configurar por medio de parámetros, en las siguientes tablas [I][II] se exponen los parámetros más relevantes de cada uno de los nodos de la red.

Tabla I
Parámetros del Nodo IMS

Parámetro	Descripción	Valor
sim_algo	Algoritmos de autenticación	Milenage
imsi	IMSI de la tarjeta SIM	00101…01
K	Clave de acceso de la SIM	0011…FF
sqn	Secuencia de información	0000…00
impi	Indica al IMSI en la red	imsi@domain
tel	Número de teléfono	0600…01

Tabla I
Parámetros del Nodo eNB

Parámetro	Descripción	Valor
dl_earfcn	Frecuencia de la portadora central para DL	3350 (2680 MHz – B7)
TDD	Modo de División Duplexación	0 (FDD)
N_RB_DL	Ancho de banda de transmisión	15 (3MHz)
N_AN_DL	Número de antenas en DL	1 (SISO)
N_AN_UL	Número de antenas en UL	1 (SISO)

V. Pruebas de Concepto

Una vez configurada e integrada correctamente, la red nos permite operar con ella. Las pruebas realizadas hasta el momento consisten en pruebas de conectividad y funcionalidad de la red 4G, siendo estas desarrolladas en la cámara anecoica, ubicada en el CITIC, de la Universidad de Granada.

Time	Diff	RAN	CN	IMS	UE ID	IMSI	Cell	Info	Message
7:30:51.003	+0.388	S1AP							Connected to 127.0.1.100:36412
-		S1AP							127.0.1.100:36412 S1 setup request
7:30:51.004	+0.001		S1AP						127.0.1.1:37564 S1 setup request
-			S1AP						127.0.1.1:37564 S1 setup response
7:30:51.005	+0.001	S1AP							127.0.1.100:36412 S1 setup response
7:31:46.699	+55.694	RRC			1		1	CCCH	RRC connection request
-		RRC			1		1	CCCH	RRC connection setup
7:31:46.760	+0.061	RRC			1		1	DCCH	RRC connection setup complete
-		NAS			1			EMM	Attach request
-		S1AP							127.0.1.100:36412 Initial UE message
-			S1AP		100	293280000000004			127.0.1.1:37564 Initial UE message
-			NAS		100	293280000000004		EMM	Attach request
-			NAS		100	293280000000004			EPS encryption caps=0xf0 integrity caps=0x70
-			NAS		100	293280000000004			GUTI not found
-			NAS		100	293280000000004		EMM	Identity request
-			S1AP		100	293280000000004			127.0.1.1:37564 Downlink NAS transport

Fig. 5. Conexión y Autenticación del UE

Tras el primer despliegue con el software Amarisoft LTE 100 se debe comprobar que la conectividad entre los eNBs y los UEs se realiza correctamente. Para ello, debemos configurar el APN (*Access Point Name*) en el UE, lanzar la red 4G y esperar a que los dispositivos se conecten automáticamente. En este caso, podemos visualizar los mensajes enviados para la autenticación y conexión desde la interfaz web en la figura 5.

Además, se han realizado pruebas de las funcionalidades propias de la red 4G como el envío de mensajes SMS y llamadas empleando VoLTE entre los UEs. A su vez, se han tomado las primeras medidas de rendimiento de la red 4G empleando SISO.

En la figura 6, comprobamos una clara tendencia creciente respecto a la capacidad del enlace (DL y UL) cuando el ancho de banda de transmisión aumenta. Además, se aprecia la mejora considerable al emplear la red 5G NSA frente a la red 4G, al igual que ocurre con el uso de la transmisión empleando técnicas mimo 2x2 frente a SISO.

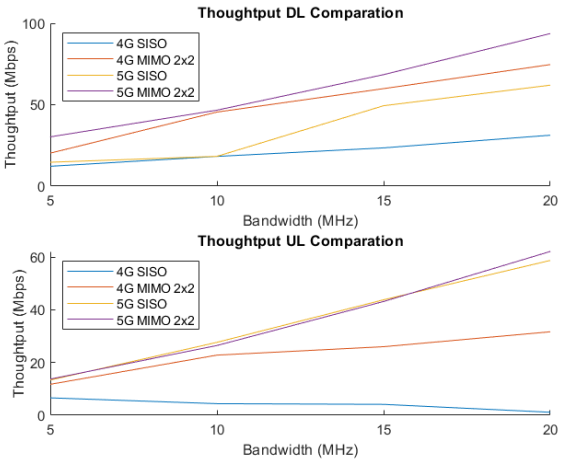


Fig. 6. Comparación de parámetros respecto ancho de banda

VI. CONCLUSIONES

Este trabajo presentaba una motivación personal y profesional debido al reto que supone el despliegue de múltiples redes y la diferenciación existente entre ellas y en la misma red empleando configuraciones diferentes.

Así, se realizó un estudio de las tecnologías disponibles para la posible implementación y se decidió optar por Amarisoft debido a las características que nos proveían. Posteriormente, se ha instalado y desplegado la red 4G que se encuentra en pleno funcionamiento y con la que se han realizado algunas de las pruebas iniciales de concepto.

VII. TRABAJOS FUTUROS

Actualmente, estamos en medio de la configuración de la estación base para el despliegue de la red 5G NSA y se pretende realizar una comparación de rendimiento respecto a la red 4G. Posteriormente, se pretende analizar la red 5G NSA mediante diferentes configuraciones para comprobar cual se adecua más a los servicios que pretende proporcionar (eMBB, mMTC, URLLC). Por último, se realizará un despliegue de la red 5G SA y se procederá de la misma manera que tenemos pensado para 5G NSA.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto H2020 5G-CLARITY (Grant No. 871428) y el Ministerio de Economía y Competitividad (proyecto TEC2016-76795-C6-4-R).

REFERENCIAS

[1] Official Page 3GPP. url: https://www.3gpp.org/ (visitado 28-04-2021).
[2] srsLTE: srsran/srsRAN. original-date: 2013-12-06T13:53:04Z. Mayo de 2021. url: https://github.com/srsran/srsRAN (visitado 21-05-2021).
[3] OpenAirInterface – 5G software alliance for democratising wireless innovation. url: https://openairinterface.org/ (visitado 21-05-2021).
[4] Official Page free5GC. url: https://www.free5gc.org/ (visitado 21-05-2021).
[5] Official Page Amarisoft: Software company dedicated to 4G LTE and 5G NR. url: https://www.amarisoft.com/ (visitado 21-05-2021).
[6] Official Page 3GPP: The Evolved Packet Core. url: https://www.3gpp.org/technologies/keywords-acronyms/100-the-evolved-packet-core (visitado 21-05-2021).
[7] "Functional architecture for the implementation of Mobile IPTV over LTE and LTE-Avanced networks", Diego F. Rueda & Zoila Ramos. December 2012 ITECKNE Innovación e Investigación en Ingeniería 9(2):40 – 51. DOI:10.15332/iteckne.v9i2.81
[8] "5G Edge Cloud Networking and Case Analysis", Lyu Huazhang, Zhang Zhonghao & Gao Shuai. October 2019. 2019 IEEE 19th International Conference on Communication Technology (ICCT), pags. 617-621. DOI: 10.1109/ICCT46805.2019.8947289 ISSN: 2576-7828
[9] Arquitectura de las redes 5G. Núcleo de red, redes RAN y arquitectura de seguridad para la tecnología 5G. Last Modified: 2021-03-19. Mayo de 2019. url: https://www.viavisolutions.com/es-es/arquitectura-de-la-tecnologia-5g (visitado 21-05-2021)


XV Jornadas de Ingeniería Telemática.
JITEL 2021.
Universidad de A Coruña.

Actas de las XV Jornadas de Ingeniería Telemática (JITEL 2021), A Coruña (España), 27-29 de octubre de 2021.

Hybrid Autonomous Connected Vehicle platooning with Federated Learning: State of the art and simulation Framework

Prashanth Kannan, Yaqoob Al-Zuhairi, Mónica Aguilar Igartua
Department of Network Engineering
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
prasuka4@gmail.com, yaqoob.al-zuhairi@upc.edu, monica.aguilar@upc.edu

In this paper, we present a brief overview about technologies and state of art methods being used for Hybrid Autonomous Connected Vehicles (HACV). Moreover, Federated Learning (FL) help in avoiding transmission of raw local data in the design of machine learning models for diverse purposes, which help to ensure privacy of sensible data. On the other hand, to reduce tailpipe emissions hybrid electric vehicles are required, as complete conversion of engines to electric might take time. Communication and connectivity of vehicles and infrastructure are increasing for connected autonomous vehicles (CAV) that play a vital role in future transportation. High speed, reliable and efficient communication between vehicles and infrastructure is made possible with fifth generation (5G) wireless technology. Here, Platooning and 5G helps in joining a cluster of vehicles that aids in reducing fuel consumption and allowing exchange of energy using wireless charging.

Keywords: Platooning, FL, VANET, V2X, Autonomous vehicles

1. INTRODUCTION

The purpose/goal of this article is to address the gap in utilizing diverse wireless technologies for vehicular networks and creating a new federated learning architecture for protection of data. In this scenario platooning of vehicles is integrated with vehicular communication technology using either 5G/Advance long term evolution (LTE) or IEEE 802.11p. This framework might help to increase safety and efficiency of vehicles and allows wireless charging of vehicle batteries. We also focus on analyzing the impact of communication failure to platoon vehicles in case of mixed traffic conditions having manned and unmanned vehicles together. The rest of the paper provides a brief outlay of current technology embedded with hybrid vehicles (fuel and electric), connected vehicles (5G and/or IEEE 802.11p), vehicular communication and federated learning (FL), along with different simulators that are useful in developing the scenarios and test our proposals.

1.1. Shared Autonomous and Connected Hybrid Vehicles

One of the current major directions of research is intelligent transportation systems (ITS) where cities are connected through autonomous vehicles (AV) and wireless technology that helps in traffic flow management, vehicular communication and decreasing CO_2 emissions to avoid the 'urban road flow congestion', reduce 'environmental pollution' and ensure 'safety and ride comfort' of the vehicles. The level of autonomy has been divided into six levels based on Society of Automotive Engineers (SAE J3016) standards, ranging from Level 0 to Level 5 [1]. According to the survey [2], 69% of respondents stated that fully automated driving will reach a 50% market share before 2050. However, understanding preferences and public response about AVs could result in Shared Autonomous Vehicles (SAV) and Connected Autonomous Vehicles (CAV) that are closely linked to ITS. SAV and CAV are beneficial when compared to that of AV [3]. It seems clear that SAVs are an interesting alternate mode of transport that replaces private vehicles [4]. However, uncertainties of utilizing SAV and CAV with respect to real world traffic made it a challenging aspect when considering mixed traffic scenarios. Also, the powertrain aspect is to be considered as a major area of research that needs to be focused, SAV and CAV with hybrid electric mode have a great potential in the depletion of tailpipe emissions and energy utilization [5]. At present, zero emission from vehicles is possible to a limited extent as complete conversion of conventional internal combustion vehicle to electric vehicle might take a lot of time due to lack of charging infrastructure, range of electric vehicle, charging time and also initial cost of the vehicle [6]. Due to the stringent emission norms and also with the strict regulations made a change to look for alternate

power sources. By 2050, European Union aims to reduce greenhouse gas emissions to zero that helps in development of research activities [7]. Hence, to reduce the percentage of CO_2 emissions hybrid electric autonomous vehicles (HEAV) play a vital role. This reduces fuel consumption by 40-50% by increasing the efficiency [8].

1.2. Platoon vehicles

Communication and collaboration of vehicles while in travel is possible using the platoon system [9, 10] which improves vehicle safety by reducing accidents, improving fuel efficiency and increasing driver comfort. Here a group of vehicles platoon by having a platoon leader (PL) vehicle that collects the information of location and road data ahead and leads the platoon members (PM) vehicles that are connected behind to navigate on road. As a result, lateral and longitudinal motion control of PM vehicles could be maintained [11]. However, in critical situations drivers should stay alert to avoid the risk of fatalities while vehicles are connected in platoon [12]. Hence, autonomous platoon vehicles (APV) are playing a vital role to eliminate the human errors when travelling in a mixed traffic flow scenario. This is operated with the help of cooperative adaptive cruise control (CACC) that combines adaptive cruise control (ACC) and vehicle to vehicle communication [13]. APV helps not only to connect commercial vehicles, but also helps to connect SAV by exchanging the vehicle information such as trajectory predictions, kinematic status, sensors and camera data. The length of platoon and number of platoon vehicles are very much crucial in managing safe travel in urban and highways scenarios.

1.3. V2V and V2X communication

To avoid fatality, maintain efficiency and increase the safety of SAV and CAV, vehicle to vehicle (V2V) and vehicle to infrastructure (V2X) communications are necessary. Here, road-side units (RSUs) are deployed over traffic sign boards, traffic signals, gas stations, etc., which establishes a wireless communication script [14]. Vehicular networks allow vehicles mounted with communication devices such as on-board units (OBUs) and other sensors that communicate using a dedicated short-range communication (DSRC) based on IEEE 802.11p standard or with cellular V2X (CV2X) [15]. Proper routing protocols developed for vehicular communication must cope with potentially high speeds of the nodes (i.e., vehicles) as well as with variable network connectivity having either sparse and dense scenarios throughout time. Also, they should consider different data priorities, e.g., sending high priority data to avoid collisions, low priority data about traffic congestion information for traffic management purposes [16].

1.4. LTE/5G in Vehicles

Advanced wireless technologies such as LTE and fifth generation (5G) communication plays a vital role in transforming V2V and V2X communication systems. [17] suggests that to ensure the safety of the vehicle, decision making and controlling need to be maintained. Implementing 5G technology along with V2X could sustain error-free navigation and accurate trajectory range than that of 4G or DSRC wireless technologies. For instance, applying the brakes in a panic situation using 4G might be 1.5 m for a vehicle, whereas for 5G it would be around just 2.5 cm. This provides promising performance improvement of braking with 5G technology compared to that of 4G [18]. Also, one of the important areas of research is hybrid communication using LTE/5G combined with DSRC/802.11p. The advantages and features of a hybrid cellular and DSRC technology is considered as an alternative for V2X communication for ITS. This has an ability of large coverage, better performance, and universal deployment when compared with only using LTE/5G or IEEE 802.11p [19, 20]. One of recent studies tests for advanced LTE have been conducted for 50 kmph at different load conditions where carriers implemented 10MHz and 15MHz to analyse the channel quality. The results reflected that carrier attained an average of 17 Mb/s and 30 Mb/s respectively [21].

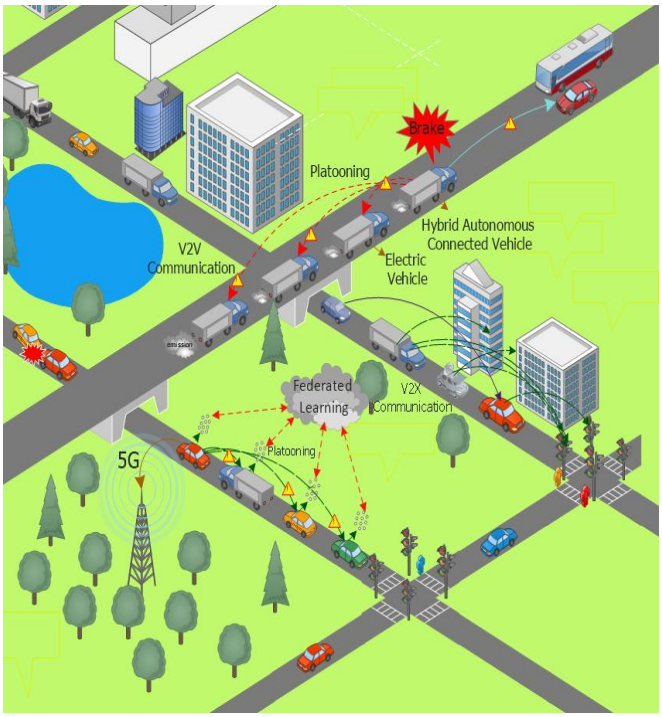


Fig.1. Typical architecture of HACV Platooning with Federated Learning.

1.5. Federated Learning in AV

With the increase in connectivity and communication of vehicles, huge amounts of data is being migrated from one point to another. Data privacy and security are major areas that need to be focused to avoid the leakage of sensible information that are being shared. Mitigation of data leakage is due to inadequate storage and computing for single node points such as vehicle or road side units. Hence, federated learning (FL) plays a vital role in managing privacy of data [22]. FL is one of the machine learning (ML) approaches where models are trained locally with data available internally. Hence, only updates from the model are being shared and direct transmission of data will not be present, thus ensuring privacy and reducing latency. The process of FL can be executed in the following process: (i) Selection of clients; (ii) Dissemination model; (iii) Distributed learning; (iv) Feedback from clients; (v) Aggregation; (vi) Testing model; (vii) Model update.

A possible implementation of FL in a vehicular network is using tree-based learning method and convolution neural

networking (CNN), which eliminates the presence of malicious data and avoids security failures to maximum extent [23]. However, as the vehicular environment is dynamic and complex, collaboration between vehicles and RSU is a challenging task that creates delay in response.

Here, we present the topics that will be studied in the first authors' doctoral thesis. In section II we highlight some of the simulators that are used to evaluate the performance of the proposed work. Section III summarizes highlights of some related work on connected autonomous and hybrid vehicles, federated learning and platooning. Since we consider the presence of hybrid autonomous electric vehicles in our proposals. Finally, section IV depicts innovative ways in collaborating the vehicles using platooning and effective ways in wireless charging of vehicles while travelling, which will be taken into account in the design of our FL framework. Fig.1. represents the basic architecture of the proposed framework platooning and wireless charging for HACV based on FL.

2. SIMULATION FRAMEWORK

To avoid real world congestion and collision with pedestrians, vehicles and other objects, AVs and CAVs are tested using multiple simulators that provide better accuracy and reliable results. Numerous simulators such as SUMO, OMNET++, VENTOS, Veins, Plexe, etc. are today available. A brief overview about main simulators used for experimentation in vehicular communications, is presented below.

2.1. SUMO

Simulation of Urban Mobility (SUMO) [24] is an open source, microscopic traffic simulator software and one of the widely used software for simulating the ITS and route planning for a particular location. SUMO also follows the physics and dynamics of various types and quantities of vehicles [24]. Generation of road network is possible with netconvert, netgenerate and netedit whereas traffic generation is possible with DUArouter, JTRrouter OD2trips, MArouter, and DFrouter. Once vehicle routing and traffic data are classified, communication between vehicles and infrastructure are classified using OMNET++.

2.2. OMNET++ and Veins

Objective Modular Network Testbed in C++ (OMNET++) [25] is a discrete event simulator with an integrated development environment (IDE) based on C++ language that helps in building network simulations. This consists of a tested and predefined framework to determine and simulate various networks, such as INET (TCP/IP stack), Veins/MIXIM (WAVE/DSRC stack), controller area network (CAN) and SimuLTE (LTE cellular network). One among them is Veins [32], which integrated with OMNET++ and SUMO to form a full simulation framework for vehicular networks. Veins includes a road and communication network that helps to simulate VANET applications for mobility vehicles.

2.3. VENTOS AND PLEXE

Vehicular Network Open Simulator (VENTOS) [26] is one of collaborative traffic flow analysis software that allows researchers to implement algorithms and protocols about DSRC. It is an open-source simulator that helps the V2X community to work with transportation, vehicular network and control theory research fields. VENTOS architecture has additional features for cooperative adaptive cruise control (CACC) in SUMO through TRACI, that helps in platooning of the vehicles. VENTOS also supports veins for V2X communication with IEEE 802.11p [27]. Another open-source simulator that is similar to VENTOS is PLEXE [28] which is mostly used for Platooning or CACC of the vehicles. As Plexe is an extension of veins framework, this helps in simulating realistic environments, maintaining the physics and dynamics of the vehicles along with an inter-vehicle communication protocol stack.

3. RELATED WORK

The following papers address key issues of the research area that need to be focused on: The work [5] implemented longitudinal autonomous driving for hybrid electric vehicles. Along with access for V2X/V2V is made simple using game theory approach for longitudinal autonomous driving control framework. This helps in achieving multiple objectives by coordinating internal combustion engine, electric motor and vehicle braking. The simulation results from the experiments suggest reduction in fuel consumption, improving driving comfort and car-following ability. The authors of [29] developed a V2V communication network to improve a truck platooning that enables sharing of data, e.g., speed of vehicle. Here, two antennas were placed at the front and rear of the PL vehicle that provides good results in a real scenario. In [30] the authors considered an advanced vehicular networking called Federated Vehicular Networking that consists of DSRC and mmWave communication for stable and scalable performance and supports machine learning (ML) and FL. Also, they have tested several routing algorithms using NetworkX, a graph-based simulator. The authors in [31] proposed a new FL algorithm for designing a CAV controller to predict and perform real time decisions. The authors proposed a novel dynamic federated proximal (DFP) algorithm for FL training. Simulation results show that the controller helps in identifying the accurate change over speed in different traffic scenarios and also, helps in developing wireless connectivity of CAVs.

4. ONGOING WORK

To improve the energy efficiency of vehicles, maintaining the safety of vehicular data and reducing the CO_2 emission are the main areas that are being focused on this thesis. The state of art (SOA) technology that is being implemented in the ongoing work is followed in four stages: (i) First stage is to design and implement a simulation framework for hybrid electric vehicle technology with shared connected autonomous vehicles that helps in maintaining the change in powertrain from engine to electric, based on the flow of traffic, type of road network such as urban area and highway. (ii) In the second stage, communication between vehicles is made for a cluster of vehicles forming a platoon. Implementing HEAV technology with CACC or platooning helps in reducing fuel consumption while vehicles are in platoon. This also helps in improving the state of charge (SOC) of the battery that helps

in charging batteries while vehicles are connected in platoons. (iii) The third stage of our ongoing work is to utilize 5G or advanced LTE combined with IEEE 802.11p/DSRC technologies that can safeguard the vehicles from collision and maintain high speeds and proper communication between vehicles without any disturbance. (iv) Finally, to sustain the privacy of data that are being shared between vehicles and RSU, advanced machine learning architectures such as FL will be employed.

5. ACKNOWLEDGEMENTS

This work was supported by the Spanish Government under research project "Enhancing Communication Protocols with Machine Learning while Protecting Sensitive Data (COMPROMISE)" PID2020-113795RB-C31/AEI/10.13039/501100011033.

REFERENCES

[1] "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On Road Motor Vehicles", *SAE Standard J3016*, 2018.

[2] M. Kyriakidis, R. Happee, J.C. de Winter, "Public opinion on automated driving: results of an international questionnaire among 5000 respondents", T*ransport Res. F: Traffic Psychol. Behav*., vol. 32, pp. 127-140, 2015.

[3] Daniela Paddeu, Ioannis Tsouros, Graham Parkhurst, Amalia Polydoropoulou, Ian Shergold, "A study of users preferences after a brief exposure in a Shared Autonomous Vehicle (SAV)", *Transportation Research Procedia*, vol. 52, pp. 533-540, 2021, https://doi.org/10.1016/j.trpro.2021.01.063.

[4] Tony Z. Zhang, T. Donna Chen, "Smart charging management for shared autonomous electric vehicle fleets: A Puget Sound case study", *Transportation Research Part D: Transport and Environment*, vol. 78, 2020, https://doi.org/10.1016/j.trd.2019.11.013.

[5] Shuo Cheng, Liang Li, Xiang Chen, Sheng-nan Fang, Xiang-yu Wang, Xiu-heng Wu, Wei-bing Li, "Longitudinal autonomous driving based on game theory for intelligent hybrid electric vehicles with connectivity", *Applied Energy*, vol. 268, 2020, https://doi.org/10.1016/j.apenergy.2020.115030.

[6] Shaik Amjad, R. Rudramoorthy, P. Sadagopan, S. Neelakrishnan, "Implementation and evaluation of change-over speed in plug-in hybrid electric two wheeler", *Energy*, vol. 109, 2016, pp. 858-865, https://doi.org/10.1016/j.energy.2016.05.025.

[7] Union, E., 2021. "EU Climate Action and the European Green Deal", Accessed 11-02-2021 2021. https://ec.europa.eu/clima/policies/strategies/2050_en.

[8] Farooq Sher, Sihong Chen, Ali Raza, Tahir Rasheed, Omid Razmkhah, Tazien Rashid, Piyya M. Rafi-ul-Shan, Begum Erten, "Novel strategies to reduce engine emissions and improve energy efficiency in hybrid vehicles", *Cleaner Engineering and Technology*, vol. 2, 2021, https://doi.org/10.1016/j.clet.2021.100074.

[9] Carl Bergenheml, Henrik Pettersson, Erik Coelingh, Cristofer Englund, Steven Shladover, Sadayuki Tsugawa, "Overview of platooning systems", *19th ITS World Congress*, Vienna, Austria, 22/26 October 2012.

[10] Heikoop, D., de Winter, J., van Arem, B., Stanton, N. A., "Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study", *Applied Ergonomics: human factors in technology and society*, vol. 60, pp. 116-127, 2017, https://doi.org/10.1016/j.apergo.2016.10.016.

[11] A. Sarker, C. Qiu, H. Shen, "Quick and autonomous platoon maintenance in vehicle dynamics for distributed vehicle platoon networks", in: *Proc. of 2nd International Conference on Internet-of-Things Design and Implementation*, pp. 203–208, 2017.

[12] Banks, V.A., Stanton, N.A., Harvey, C., "Sub-systems on the road to vehicle automation: hands and feet free but not 'mind' free driving", *Safety Science*, vol. 62, pp. 505-514, 2014, http://dx.doi.org/10.1016/j.ssci.2013.10.014.

[13] P.K. Singh, G.S. Tabjul, M. Imran, S.K. Nandi, S. Nandi, "Impact of security attacks on cooperative driving use case: CACC platooning", in: Proc. of TENCON, pp. 0138–0143, 2018.

[14] Hong Zhang, Xinxin Lu, "Vehicle communication network in intelligent transportation systems based on Internet of Things", *Computer Communications*, vol. 160, pp. 799–806, 2020.

[15] Michael Lee, Travis Atkison, "VANET applications: Past, present, and future", *Vehicular Communications*, vol. 28, 2021, https://doi.org/10.1016/j.vehcom.2020.100310.

[16] Dahiya, A., Chauhan, R., "A comparative study of MANET and VANET environment", *Journal of computing*, vol. 2, no. 7, pp. 87-92, 2010.

[17] Y. Shi, Q. Han, W. Shen, X. Wang, "A Multi-Layer Collaboration Framework for Industrial Parks with 5G Vehicle-to-Everything Networks", *Engineering*, 2021, https://doi.org/10.1016/j.eng.2020.12.021

[18] Guevara L, Auat Cheein F. "The role of 5G technologies: challenges in smart cities and intelligent transportation systems", *Sustainability*, 12(16):6469, 2020, https://doi.org/10.3390/su12166469.

[19] G. Araniti, C. Campolo, M. Condoluci, A. Iera and A. Molinaro, "LTE for vehicular networking: a survey", in *IEEE Communications Magazine*, vol. 51, No. 5, pp. 148-157, May 2013, doi: 10.1109/MCOM.2013.6515060.

[20] A. Burbano-Abril, R.M. Edwards, V. Rangel-Licea, R. Aquino-Santos, M. Lopez-Guerrero, R.S. Kalawsky, M. Behjati, "Modeling and analysis of LTE connectivity in a high mobility vehicular environment", *Computers & Electrical Engineering*, vol. 68, pp. 322-336, 2018, https://doi.org/10.1016/j.compeleceng.2018.04.007.

[21] S. Avallone, N. Pasquino, S. Zinno and D. Casillo, "Smartphone-based measurements of LTE network performance", *2017 IEEE International Instrumentation and Measurement Technology Conference* (I2MTC), pp. 1-6, 2017, doi: 10.1109/I2MTC.2017.7969854.

[22] Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, Dave Bacon, "Federated Learning: Strategies for Improving Communication Efficiency", arXiv:1610.05492, 2016.

[23] Y. Lu, X. Huang, Y. Dai, S. Maharjan and Y. Zhang, "Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems", *IEEE Network*, vol. 34, no. 3, pp. 50-56, May/June 2020, doi: 10.1109/MNET.011.1900317.

[24] Daniel Krajzewicz, 2010, "Traffic Simulation with SUMO – Simulation of Urban Mobility", International Series in Operations Research & Management Science, Fundamentals of Traffic Simulation, chapter 0, pp. 269-293, Springer, DOI: 10.1007/978-1-4419-6142-6_7.

[25] C. Sommer, R. German and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis", *IEEE Transactions on Mobile Computing*, vol. 10, no. 1, pp. 3-15, Jan. 2011, doi: 10.1109/TMC.2010.133.

[26] Amoozadeh, M., "Vehicular Network Open Simulator (VENTOS)", http://maniam.github.io/VENTOS.

[27] Mani Amoozadeh, Bryan Ching, Chen-Nee Chuah, Dipak Ghosal, H. Michael Zhang, "VENTOS: Vehicular Network Open Simulator with Hardware-in-the-Loop Support", *Procedia Computer Science*, vol. 151, 2019, pp. 61-68, https://doi.org/10.1016/j.procs.2019.04.012.

[28] M. Segata, S. Joerer, B. Bloessl, C. Sommer, F. Dressler and R. L. Cigno, "Plexe: A platooning extension for Veins", *2014 IEEE Vehicular Networking Conference (VNC)*, pp. 53-60, 2014, doi: 10.1109/VNC.2014.7013309.

[29] Carl Bergenhem, Erik Hedin, Daniel Skarin, "Vehicle-to-Vehicle Communication for a Platooning System", *Procedia - Social and Behavioral Sciences*, vol. 48, pp. 1222 – 1233, 2012.

[30] J. Posner, L. Tseng, M. Aloqaily and Y. Jararweh, "Federated Learning in Vehicular Networks: Opportunities and Solutions", *IEEE Network*, vol. 35, no. 2, pp. 152-159, 2021, doi: 10.1109/MNET.011.2000430.

[31] Tengchan Zeng, Mingzhe Chen, Mehdi Bennis, "Federated Learning on the Road: Autonomous Controller Design for Connected and Autonomous Vehicles", arXiv:2102.03401, 5 Feb 2021.

[32] Iftikhar Ahmad, Rafidah Md Noor, Ismail Ahmedy, Syed Adeel Ali Shah, Ibrar Yaqoob, Ejaz Ahmed, Muhammad Imran, "VANET–LTE based heterogeneous vehicular clustering for driving assistance and route planning applications", *Computer Networks*, vol. 145, pp. 128-140, 2018, https://doi.org/10.1016/j.comnet.2018.08.018.


citic
UNIVERSIDADE DA CORUÑA
A Coruña
SIEMPRE ABIERTA
TURISMO CORUÑA
Xacobeo 21·22
UNIÓN EUROPEA
XUNTA DE GALICIA
jitel21.udc.es