

# Marco para el Análisis e Inferencia de Conocimiento en Redes 5G

Marco Antonio Sotelo Monge, Jorge Maestre Vidal, Luis Javier García Villalba  
Grupo de Análisis, Seguridad y Sistemas (GASS)  
Departamento de Ingeniería del Software e Inteligencia Artificial (DISIA)  
Facultad de Informática, Despacho 431, Universidad Complutense de Madrid  
Calle Profesor José García Santesmases, 9, Ciudad Universitaria, 28040 Madrid, España  
[masotelo@ucm.es](mailto:masotelo@ucm.es), [jmaestre@ucm.es](mailto:jmaestre@ucm.es), [javierv@fdi.ucm.es](mailto:javierv@fdi.ucm.es)

**Resumen**—Este trabajo propone un marco de análisis de información para la inferencia de conocimiento en el contexto del proyecto SELFNET, cuya finalidad es diagnosticar el estado de la red y predecir problemas potenciales que afecten la operatividad de la red, facilitándose también el proceso de toma de decisiones en redes 5G. Esta propuesta proporciona capacidades para el descubrimiento de hechos, reconocimiento de patrones, razonamiento y predicción con el objetivo de inferir conductas sospechosas que puedan ser mitigadas a través del despliegue de medidas de respuesta, tanto de forma reactiva como proactiva. Además, este marco de análisis utiliza una metodología basada en casos de uso, donde el operador es capaz de personalizar los parámetros de operación y las reglas para la inferencia de conocimiento. La propuesta ha sido evaluada sobre un caso de uso, donde se demuestra cómo a partir de una configuración sencilla es posible abastecer a las capas de inteligencia del conocimiento necesario para mejorar la toma de decisiones que permita adaptar la red a cambios en el volumen de datos monitorizados.

**Palabras Clave**—5G, Conciencia Situacional, Inferencia de Conocimiento, NFV, Predicción, SDN.

## I. INTRODUCCIÓN

El crecimiento del número de dispositivos conectados a las infraestructuras móviles actuales y la demanda de servicios en línea han generado nuevos desafíos en la gestión de las infraestructuras de redes y telecomunicaciones. Los sistemas actuales requieren de respuestas rápidas frente a problemas típicos de red, tales como caída de enlaces o congestión, con el objetivo de garantizar la Calidad de Servicio (QoS) y la Calidad de Experiencia (QoE) de los usuarios finales. Se busca también que el tiempo de recuperación de servicios y los costes de capital (capex) y de operación (opex) disminuyan [2] cada vez más. Actualmente, la personalización de los servicios de red se ve limitada por la rigidez de las arquitecturas tradicionales debido a que tienen una fuerte dependencia de la configuración manual de dispositivos, así como de un lento proceso de estandarización (desde su diseño hasta su

implementación) para la generación de nuevos productos y servicios. Según el reporte de la comisión Europea “Future Internet” la próxima generación de redes móviles (5G) superará dichas limitaciones, y se pronostican sus primeros resultados para el año 2020 [1]. Éstas tienen por objetivo el proporcionar un entorno de comunicación confiable y de alto rendimiento que asegure la provisión eficiente de servicios, y que a su vez garantice el cumplimiento de los Acuerdos de Nivel de Servicio (SLAs) [3]. La nueva generación se caracterizará por indicadores de desempeño muy superiores a los ofrecidos por las generaciones predecesoras, considerando entre ellos las altas capacidades de transmisión, baja latencia, altos niveles de QoE/QoS (evaluados según métricas objetivas [4]), gran densidad de dispositivos por área geográfica, etc. Por lo tanto, 5G debe ser capaz no sólo de responder automáticamente a situaciones imprevistas que comprometan la operatividad de la red, sino también debe ofrecer un modelo de control heterogéneo y unificado [5]. Para alcanzar los objetivos planteados, 5G propone la integración de diversas tecnologías como Redes Definidas por Software (SDN) [6], Virtualización de Funciones de Red (NFV) [7], computación en la nube [8], Inteligencia Artificial (IA) [9], Redes Auto-Organizadas (SON) [10], entre otras. La combinación de dichas tecnologías facilitará el despliegue y gestión de nuevos servicios en un entorno abierto y con capacidades mejoradas en la utilización del espectro, virtualización y compartición de recursos, eficiencia energética, etc. SDN ha emergido como una nueva arquitectura en la cual el control de los dispositivos de red está centralizado en el controlador SDN, gestionando el comportamiento de la red a través de aplicaciones SDN. Asimismo, diversas funciones virtualizadas de red (NFVs) podrían ser desplegadas automáticamente y a demanda desde el controlador en un entorno virtualizado. El análisis del tráfico que circula a través de la red permitirá por lo tanto recurrir a esquemas de análisis más complejos que

harán posible inferir síntomas que proyecten escenarios de red que afecten el cumplimiento de los indicadores de desempeño en una red 5G, y den lugar a procesos complejos de toma de decisiones que se traduzcan en medidas de actuación. Con este objetivo, el presente trabajo propone la creación de un Marco para el Análisis e Inferencia de Conocimiento en Redes 5G que cumpla los requerimientos antes mencionados.

El trabajo se estructura en cinco secciones, siendo esta introducción la primera. La sección II describe los trabajos relacionados y el proyecto SELFNET. En la sección III se presenta el modelo de análisis para la inferencia de conocimiento en 5G. La sección IV describe los resultados preliminares del modelo propuesto. Finalmente, en la sección V se presentan las conclusiones.

## II. TRABAJOS RELACIONADOS

La comunidad científica, académica, operadores de telecomunicaciones, proveedores de servicio, entre otros, han unido esfuerzos para el desarrollo de soluciones que mejoren la gestión y el análisis de información en redes 5G, tomando las ventajas de SDN, NFV, SON e inteligencia artificial. Tal es el caso del consorcio 5G Americas [11], o las iniciativas para la adopción de 5G en Asia [12], entre otras. En Europa, el consorcio 5G-PPP [13] promueve el desarrollo de proyectos de investigación que brinden soluciones a los desafíos de 5G a través de la sinergia de diversas áreas del conocimiento. En la Tabla I se describen proyectos relevantes en el ámbito de 5G.

En particular, el proyecto SELFNET (Self-Organized Network Management in Virtualized and Software Defined Networks) [14] ha sido creado con el fin de proporcionar un marco de gestión autónomo que provea capacidades de auto-organización en redes móviles 5G. Este proyecto se encuentra en desarrollo y es financiado por el programa marco Horizonte 2020. SELFNET hace uso de los principios de SDN y NFV para gestionar de forma inteligente y autónoma diversas funciones de red con el propósito de dar solución automática a problemas comunes de red, tales como el retardo en la transmisión, la caída de enlaces, etc. Para ello, SELFNET [15] integra un paradigma de autogestión de la red basado en el uso de algoritmos de predicción, reconocimiento de patrones, técnicas de minería de datos, entre otros, para identificar el comportamiento de una red móvil 5G, y determinar las mejores acciones que mitiguen de forma automática los problemas de red detectados. SELFNET se basa en la definición de tres grandes casos de uso: la autoprotección (self-protection), la autosalud (self-healing) y la autooptimización (self-optimization), en los que se distinguen escenarios sobre los que el sistema requiera procesos de toma de decisiones y el despliegue de acciones de respuesta. Para este propósito, un conjunto de sensores y actuadores SDN/NFV (por ejemplo, Sistemas de Detección de Intrusión (IDS), servidores de Inspección Profunda de Paquetes (DPI), anti-malware, etc.) serán desplegados para analizar el estado de la red y detectar situaciones que sean controladas de forma reactiva o proactiva. Los

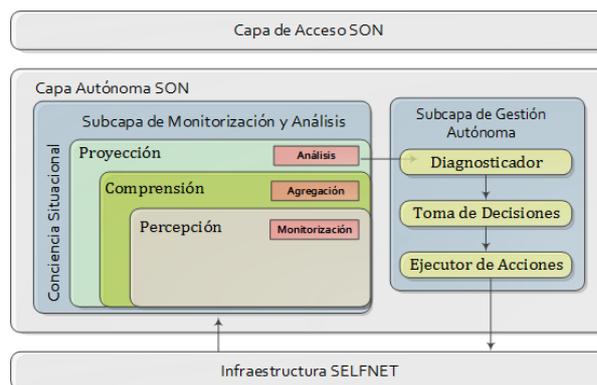


Fig. 1. Modelo de Conciencia Situacional en SELFNET.

sensores y actuadores pueden ser desplegados tanto en la infraestructura física como virtual de SELFNET.

En SELFNET la inteligencia de red es provista por la capa Autónoma SON. Esta capa recopila métricas relacionadas con el comportamiento de la red y utiliza esta información para inferir el estado de la misma. Luego, decide las acciones a ser ejecutadas para cumplir los objetivos del sistema. En particular, SELFNET proporciona un sistema de Monitorización, Agregación y Correlación, y Análisis (Subcapa de Monitorización y Análisis), que facilita el conocimiento del contexto de operación a través del modelo de consciencia situacional propuesto por Endsley [16] (ver Fig. 1). Los procesos de SELFNET previos a la toma de decisiones son descritos a continuación:

- **Monitorización.** El principal objetivo de la monitorización es recopilar información proveniente de la infraestructura de red física y virtual, en forma de métricas de bajo nivel y de eventos, para facilitar su acceso a capas superiores.
- **Agregación y correlación.** En esta etapa las métricas de bajo nivel sirven de entrada a procesos de agregación con el objetivo de reducir el alto volumen de datos recibido, obteniéndose métricas agregadas sobre algún dominio particular del sistema. Por otra parte, los eventos recibidos son correlacionados con el fin de filtrar información no sensitiva o redundante (por ejemplo, eliminando alertas repetidas) y también de obtener una visión global de la red en función del contexto de operación.
- **Análisis.** En el componente de análisis se lleva a cabo la identificación de escenarios que potencialmente amenacen la operatividad de la red a partir de la información sensada. Cuando estos escenarios son inferidos son enviados a la Subcapa de Gestión Autónoma, la cual utiliza diferentes algoritmos y técnicas avanzadas de inteligencia para determinar la causa del problema (Diagnosticador). Luego se define la mejor estrategia de reacción (Toma de Decisiones), para que finalmente se lleven a cabo las acciones de respuesta o mitigación del problema (Ejecutor de Acciones).

Este artículo se centra en el módulo de análisis y la inferencia de conocimiento.

Tabla I  
PROYECTOS 5G.

Proyecto	Tecnología	Descripción
CROWD [17]	SDN, NFV	Este proyecto tiene el objetivo de aumentar la capacidad en el parámetro de densidad de las redes de acceso inalámbrico en redes heterogéneas. Se enfoca en garantizar la QoE de los usuarios móviles, la optimización de los recursos y el consumo de energía.
T-NOVA [18]	SDN, NFV	Este proyecto se enfoca en el despliegue de Funciones de Red como Servicio (NFaaS) sobre infraestructuras de red virtualizadas. T-Nova diseña e implementa una plataforma de gestión y orquestación automatizada para la provisión, configuración, monitorización y optimización de recursos virtualizados, basada en SDN y NFV.
UNIFY [19]	SDN, NFV	El proyecto tiene como objetivo desarrollar una plataforma dinámica para la creación de servicios. Unify permite el despliegue dinámico y automático de los servicios en entornos cloud (red, computación y almacenamiento). De manera similar, el orquestador incluye algoritmos de optimización para la asignación óptima de dichos recursos a lo largo de la infraestructura.
COGNET [20]	SDN, NFV, Aprendizaje de máquina	Este proyecto se centra en el mejoramiento de las tareas de monitorización y la gestión automática de red, para lo cual predice la demanda de recursos y luego cambia su configuración basada en el análisis actual de la red (detección de errores, fallos de seguridad).
SELFNET [15]	SDN, NFV	El proyecto provee un marco de gestión autónoma de red, basado en el despliegue inteligente de sensores y actuadores tanto en la infraestructura física como virtual. Las funciones pueden ser basadas en SDN o NFV.
CHARISMA [21]	SDN, NFV	Este proyecto permite el despliegue inteligente de servicios para redes de acceso de radio en entornos cloud (C-RAN).
5G-ENSURE [22]	SDN, NFV	El proyecto se enfoca en la evaluación de riesgos y su mitigación para cubrir los requerimientos de seguridad de 5G. 5G-Ensure cubre un rango amplio de problemas de seguridad, desde el aseguramiento de dispositivos físicos hasta de recursos SDN o NFV.

### III. MARCO PARA EL ANÁLISIS E INFERENCIA DE CONOCIMIENTO EN 5G

El componente de análisis tiene el objetivo de inferir conocimiento sobre la base de la información proveniente de los niveles inferiores de la arquitectura (monitorización, agregación y correlación), para identificar situaciones potenciales (síntomas) que desencadenen procesos avanzados de diagnóstico y toma de decisiones en el componente de inteligencia de SELFNET (ver Fig. 2). A continuación se describen cada uno de sus componentes y su relación con el resto de elementos del sistema.

#### A. Componentes

Se distinguen siete pasos principales para el procesamiento de información: Integración de Casos de Uso [O], Descubrimiento [DIS], Reconocimiento de Patrones [PR], Predicción [FT], Umbrales Adaptativos [ATh], Inferencia de Conocimiento [KI] y Notificación [N].

- *Integración de Casos de Uso*. En este conjunto de acciones se definen los descriptores que permiten la elaboración de la ontología de información, tanto a nivel procedimental como factual, necesaria para configurar el marco de análisis en función de los requisitos de los casos de uso. A partir de ellos se definen desde los objetos  $O$  a tener en cuenta, hasta la reglas  $Ru$  de inferencia de conocimiento.
- *Descubrimiento [DIS]*. Esta operación permite enlazar las capas de agregación y de análisis de SELFNET para que la primera envíe periódicamente datos agregados construidos a partir de las observaciones extraídas por los sensores desplegados a lo largo del escenario de monitorización. El módulo de análisis traduce dichas observaciones en hechos ( $Fa$ ) que se añaden a la memoria de trabajo, para que en lo posterior se lleven a cabo las tareas de reconocimiento de patrones, predicción o umbrales adaptativos.

- *Reconocimiento de Patrones [PR]*. En este proceso se identifican ciertos patrones ( $PR$ ) adquiridos o conocidos anteriormente sobre hechos ( $Fa$ ) relacionados con datos agregados o eventos, y se generan nuevos hechos ( $Fa(PR)$ ) derivados del conocimiento resultante de su estudio. Con este propósito, diferentes tareas internas son ejecutadas: estudio de los datos de entrada (tanto los datos de entrenamiento como las observaciones realizadas), decisión de la estrategia de datos más adecuada para cada contexto, características de extracción, construcción de modelos o regresiones, etc. La selección de métodos de reconocimiento de patrones es adaptada a las necesidades de cada escenario de seguridad. El marco de análisis de SELFNET se enfoca en dos acciones fundamentales: la identificación de firmas de eventos conocidos previamente y la detección de anomalías.
- *Predicción [FT]*. En esta etapa se lleva a cabo el cálculo de las métricas de predicción ( $Ft$ ), expresadas como hechos ( $Fa(Ft)$ ), asociadas a cada escenario a partir de las observaciones provistas por la fase de agregación y correlación. Este proceso implica diferentes etapas: gestión del historial de datos requeridos para construir el modelo de predicción, decisión del algoritmo de predicción más adecuado, y la evaluación de los resultados que faciliten el aprendizaje basado en las decisiones previas. El pronóstico de situaciones de red facilita la optimización de recursos, el despliegue de acciones proactivas y permite anticipar la identificación de riesgos en el entorno de monitorización. Este componente incluye una batería de algoritmos de predicción (alisamiento exponencial, medias móviles, modelo autorregresivo integrado de media móvil (ARIMA), etc.) que son evaluados con distintos parámetros de ajuste para determinar el método de predicción aplicado a la serie temporal que

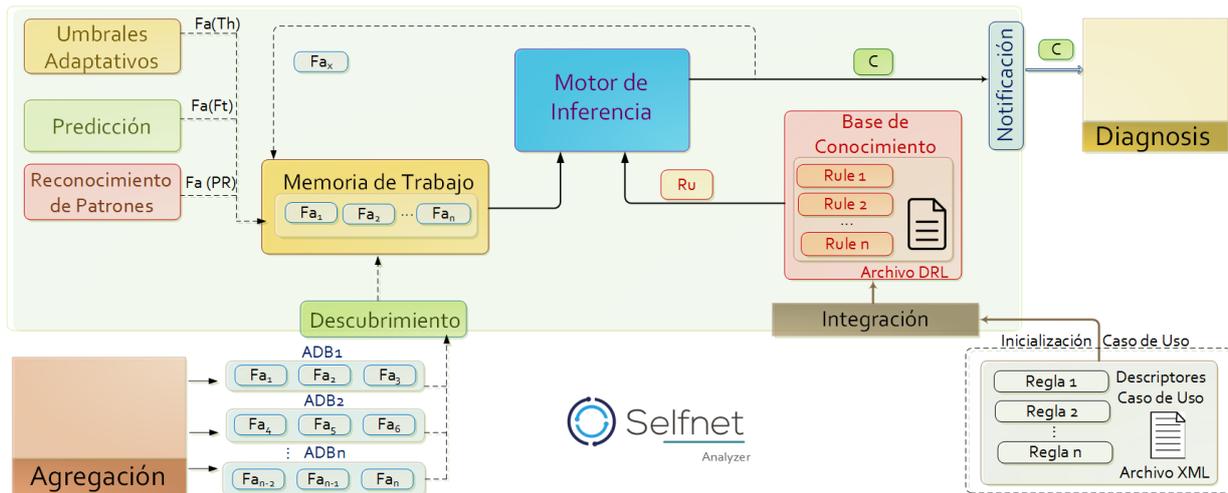


Fig. 2. Inferencia de Conocimiento en SELFNET

minimiza el error de estimación. Con el método seleccionado, el marco de análisis de SELFNET predice los horizontes requeridos para una serie temporal dada.

- **Umbrales adaptativos [ATH].** En este proceso se llevan a cabo las acciones que permiten definir cuando los errores de predicción deben ser tenidos en cuenta. De este modo se construyen umbrales ( $Th$ ) y se genera nuevos hechos ( $Fa(Th)$ ) a partir de ellos. Su elaboración implica diferentes pasos, tales como el análisis y la extracción de las características principales de los datos de entrada, decisión de los algoritmos más adecuados, o modelado y estimación de los valores en los umbrales. La principal aplicabilidad de los umbrales adaptativos en SELFNET es que considera el contexto de operación en la inferencia de nuevos hechos relacionados con el filtrado, reduciendo así las tasas de falsos positivos de sus sensores.
- **Inferencia de Conocimiento [KI].** En esta tarea se aplican las reglas ( $Ru$ ) de producción contenidas en la base de conocimiento con el objetivo de deducir nuevo conocimiento. Las reglas son generadas por el operador del sistema, y se determinan por cada caso de uso. El motor de inferencia opera bajo un esquema de encadenamiento hacia adelante (*forward chaining*), es decir, considera primero los hechos conocidos con anterioridad e infiere nuevos hechos hasta que sea capaz de inferir conclusiones ( $C$ ) [25], [26]. Además, es importante tener en cuenta que la implementación más sencilla del motor de inferencia considera reglas de separación (modus ponens) basadas en lógica proposicional. Las reglas pueden ser adaptadas a una representación diferente de incertidumbre, tal como la lógica difusa, conjuntos aproximados, o redes Bayesianas.
- **Notificación [N].** Esta última etapa tiene por objetivo actuar como enlace entre el conocimiento adquirido por el módulo de Análisis y la subcapa de inteligencia de SELFNET (Subcapa Autónoma SON).

Adicionalmente, el marco de análisis provee un módulo de Interfaz de Usuario que permite al operador configurar las reglas de inferencia por cada caso de uso, así como un módulo de Notificación cuya función es enviar las conclusiones obtenidas al componente de inteligencia.

### B. Entradas y salidas

Desde el componente de agregación llegan métricas agregadas y eventos correlacionados. Los módulos internos del marco de análisis (reconocimiento de patrones, predicción y umbrales adaptativos) actúan como proveedores internos de nuevos hechos que sirven como entradas en la memoria de la base de conocimiento. Asimismo, el conjunto de reglas para un caso de uso en particular, desde la interfaz de usuario, complementa el conjunto de entradas necesarias. Las conclusiones del proceso de inferencia de conocimiento son expresadas como reportes que se envían como salida al componente de inteligencia de SELFNET.

## IV. EXPERIMENTACIÓN Y RESULTADOS PRELIMINARES

### A. Experimentación

La experimentación se conduce sobre un caso de uso de ejemplo, implementado para la obtención de resultados. El objetivo es recorrer las etapas definidas en el marco de análisis para la generación de conocimiento; las cuales incluyen un mecanismo básico de reconocimiento de patrones, predicción de datos sobre una serie temporal, construcción de intervalos de predicción adaptativos, y aplicación de reglas de producción con el fin de generar conclusiones que denoten el estado de la red.

#### 1) Conjuntos de Muestras:

Se describen a continuación los dos conjuntos de muestras usados en la experimentación. El primero, con fines de validación del componente de predicción. El segundo, para la generación de conocimiento (conclusiones) en el motor de inferencia.

#### Competición M3

Para evaluar la precisión del componente de predicción se

Tabla II  
CARACTERÍSTICAS DE LAS SERIES TEMPORALES EN LA  
COMPETICIÓN M3.

Frecuencia	Series Temporales	Observaciones	Horizontes
Anual	645	19	6
Cuatrimestral	756	44	8
Mensual	1428	115	18
Otra	174	63	8

cuenta con un conjunto de muestras de propósito general usado en la competición M3 [23]. Esta competición reúne un conjunto de 3003 series temporales sobre las que se aplicaron diversos métodos de predicción en distintos horizontes ( $t + 1$ ,  $t + 2$ , ...,  $t + 18$ ), según el número de observaciones contenidas en ellas. Las características y horizontes evaluados en cada caso se resumen en la Tabla II. La precisión de los métodos de predicción aplicados en esta prueba se compara luego con los resultados obtenidos por SELFNET.

#### *Tráfico de red UCM 2011*

Se cuenta además con un conjunto de muestras de tráfico real elaborado a partir de capturas proporcionadas por el Centro de Cálculo y Procesamiento de Datos de la Universidad Complutense de Madrid (UCM). Esta colección consta de diferentes trazas de tráfico monitorizado en la facultad de informática de la UCM durante varios días, en diferentes periodos de tiempo, a lo largo del año 2011 [27]. En la experimentación realizada han sido compactadas en formato pcap [27], y son usadas para predecir el volumen de tráfico de red en los siguientes horizontes temporales. Por cada paquete contenido en el archivo pcap se obtiene como métrica el número de bytes transmitidos. Este valor corresponde a una métrica de bajo nivel en la capa de monitorización de SELFNET. El número de bytes transmitidos se acumula (mediante un promedio) en intervalos de 5 segundos generando así una serie temporal en la que cada observación promediada, expresada en Kilobytes, representa una métrica agregada. Esta operación corresponde a la capa de agregación de SELFNET y la serie temporal producida sirve como información de entrada para el marco de análisis propuesto.

2) *Caso de Uso: Comportamiento anómalo del volumen de tráfico:* El objetivo de este caso de uso es inferir si el volumen de tráfico de la red presenta un comportamiento anómalo. Los distintos componentes del marco de análisis contribuyen a la generación de hechos en la memoria de trabajo, sobre los que el motor de inferencia aplicará las reglas de producción configuradas. El proceso se inicia cuando se recibe la serie temporal del volumen de tráfico. Cada vez que se añade un nuevo elemento a la serie, el componente de reconocimiento de patrones (PR) evalúa si existe una tendencia creciente o decreciente en las observaciones, e introduce nuevos hechos en la memoria de trabajo que registren dicho patrón. En la siguiente etapa de procesamiento de información se lleva a cabo su predicción, estableciéndose un horizonte temporal de 8 observaciones para la generación de las estimaciones que serán insertadas en la memoria de trabajo. Una vez conoci-

das las métricas pronosticadas, se definen los intervalos de predicción (PI) para cada nueva observación que se añade a la serie temporal, y se registran dichos intervalos en la memoria de trabajo del sistema experto. El PI se compone de un umbral superior (U. Sup) y uno inferior (U. inf) que se generan a través de la comparación entre el valor observado y el valor pronosticado, según se detalla en [24]. El patrón identificado, las predicciones, y los umbrales, permiten al motor de inferencia determinar si el tráfico de red es anómalo cuando éste se encuentra fuera del PI. Además, el motor de inferencia evalúa el patrón de la observación para conocer si el volumen de tráfico anómalo se acerca o aleja del intervalo de predicción (congestión o subutilización de la red, respectivamente). Por ejemplo, si una observación ha excedido el umbral superior y su tendencia es creciente, se generará un síntoma de congestión con tendencia positiva (+), lo que podría desencadenar el despliegue de acciones de mitigación específicas desde el componente de inteligencia de SELFNET, tales como la ampliación del ancho de banda o la instanciación de nuevas funciones de red; por el contrario, si el síntoma determina una tendencia negativa (-), es posible que se decida eliminar instancias de funciones de red a fin de mejorar la Calidad de Servicio (QoS) y la Calidad de Experiencia (QoE), así como de reducir el consumo de recursos.

#### *B. Resultados*

Los resultados se han dividido en dos secciones. En la primera, se validaron las capacidades de predicción del marco propuesto. En la segunda sección, se presentan los resultados obtenidos tras la implementación del caso de uso.

##### *1) Evaluación del módulo de Predicción:*

La precisión de los resultados obtenidos por el componente de predicción ha sido comparada con la precisión obtenida por diversos métodos de predicción aplicados en la competición M3. La precisión fue evaluada mediante el cálculo del Error Porcentual Absoluto Medio Simétrico (SMAPE) [23], cuyo valor fluctúa entre 0 y 200%, siendo éste el error de predicción en cada caso. Dado que los errores se calcularon en varias series temporales (por ejemplo 756 en datos cuatrimestrales), se obtuvo el valor promedio del SMAPE por cada horizonte de predicción y por cada método aplicado. Nótese que cuanto menor es el valor del error SMAPE promedio, más precisa es la predicción que evalúa. En la experimentación se comparó el SMAPE promedio obtenido por SELFNET con el mínimo valor de SMAPE promedio hallado en la competición M3 para un determinado horizonte temporal, encontrándose que SELFNET obtuvo, en general, mejores resultados de predicción en todas las series temporales analizadas y en casi todos los horizontes evaluados. En las 645 series temporales anuales, SELFNET mostró un mejor resultado en todos los horizontes temporales. Estos resultados se resumen en la Tabla III.

En las 756 series cuatrimestrales, SELFNET fue superado sólo en el horizonte  $t+1$  por el método de predicción *PP-autocast*, cuyo SMAPE promedio fue de 4.8, siendo

Tabla III  
VALORES SMAPE DE M3 Y SELFNET EN LAS SERIES ANUALES

Método	1	2	3	4	5	6
M3 mejor	7.6	11.8	16.1	18.2	13.4	22.7
SELFNET	6.9	6.6	7.6	7.2	8.5	9.4

Tabla IV  
VALORES SMAPE DE M3 Y SELFNET EN LAS SERIES CUATRIMESTRALES

Método	1	2	3	4	5	6	8
M3 mejor	4.8	6.6	7.4	8.8	9.4	10.9	12
SELFNET	5.3	5.2	4.5	4.7	4.4	4.8	4.9

inferior al 5.3 que obtuvo SELFNET. En todos los demás casos, la precisión de SELFNET fue superior. Estos resultados se resumen en la Tabla IV.

En las 1428 series mensuales, SELFNET fue superado por M3 sólo en dos ocasiones. En el horizonte t+2, el método ForecastPro obtuvo un SMAPE promedio de 10.7 (menor a SELFNET en 0.5), mientras que en el horizonte t+4 el método Theta obtuvo un SMAPE promedio de 12.4, inferior al de SELFNET en 0.1. En todos los demás casos, los resultados de SELFNET fueron mejores. Estos resultados se resumen en la Tabla V.

En las otras 174 series temporales, SELFNET mostró un mejor resultado en todos los horizontes temporales, con excepción de t+1 en donde el método Autobox 2 de M3 obtuvo un SMAPE promedio de 1.6 que fue inferior a SELFNET en 0.2. Estos resultados se resumen en la Tabla VI.

2) *Evaluación del caso de uso:* La serie temporal Tráfico de red UCM 2011 cuenta con 31 elementos. Por ello, se hicieron las predicciones para 8 horizontes temporales debido a que el tamaño de esta serie temporal es cercano al tamaño promedio de las series cuatrimestrales en la competición M3. No obstante, este parámetro es configurable en el marco de análisis de SELFNET. La detección del patrón respondió a un algoritmo básico que comparó la observación actual con la anterior para determinar si existe una tendencia creciente o decreciente en los datos de la muestra. Los patrones identificados se muestran en la Tabla VII.

Para hallar las predicciones se restaron ocho elementos a la serie temporal original, los que fueron luego pronosticados y comparados con los valores originales para estimar la precisión. Para el tráfico de red UCM, el componente de predicción de SELFNET seleccionó el algoritmo de Suavizamiento Exponencial Simple (SES), con  $\alpha$  aproximado a 0.7 y un SMAPE medio de 5.342725 (obtenido de los ocho horizontes), como el algoritmo

Tabla VI  
VALORES SMAPE DE M3 Y SELFNET EN LAS OTRAS SERIES TEMPORALES

Método	1	2	3	4	5	6	8
M3 mejor	1.6	2.7	3.8	4.3	5.3	5.1	6
SELFNET	1.8	2.3	2.2	2	2.3	1.5	2.4

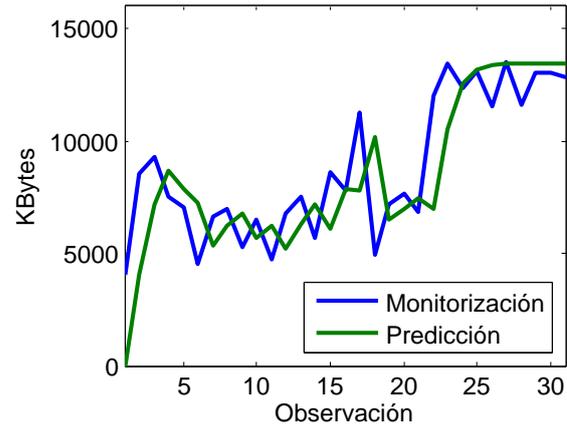


Fig. 3. Serie Temporal Real y Serie Pronosticada para UCM 2011.

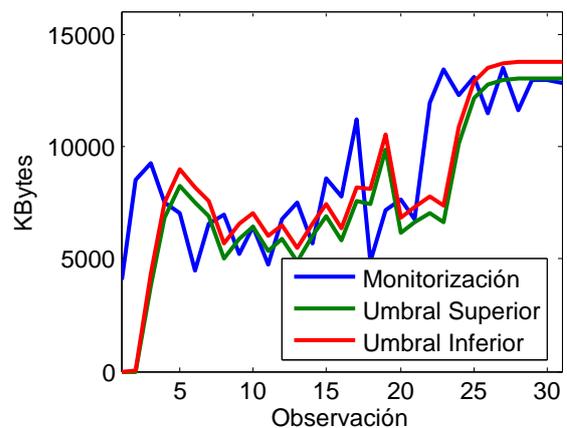


Fig. 4. Observaciones y Umbrales Calculados para UCM 2011.

que mejores resultados obtuvo tras la comparativa con otros métodos de predicción contenidos en su batería de algoritmos. Los resultados se muestran en la Tabla VII. A su vez, la serie original y la serie pronosticada se grafican en Fig. 3.

A partir de los resultados pronosticados SELFNET estimó el intervalo de predicción, construyendo para ello los umbrales adaptativos superior e inferior con un valor

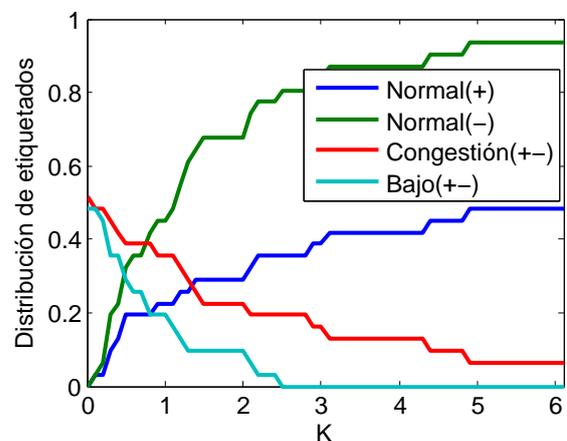


Fig. 5. Resultados al variar K para UCM 2011.

Tabla V  
VALORES SMAPE DE M3 Y SELFNET EN LAS SERIES ANUALES

Método	1	2	3	4	5	6	8	12	15	18
M3 mejor	11.2	10.7	11.7	12.4	11.8	12.2	12.6	13.2	16.2	17.8
SELFNET	11	11.2	11.7	12.5	11.6	11.4	10.6	9.6	11	12.7

Tabla VII  
RESULTADOS OBTENIDOS POR EL MARCO DE ANÁLISIS EN EL CASO DE USO EVALUADO

N	KPI(Volumen) (KB)	Patrón	Predicción	U. Sup.	U. Inf.	Síntoma
1	4063	Crece	0.0	0.0	0.0	Congestión (+)
2	8557	Crece	4062.5	215.8	-215.8	Congestión (+)
3	9264	Crece	7208.4	5125.3	2999.8	Congestión (+)
4	7498	Decrece	8647.0	8592.8	5824.0	Normal (-)
5	7030	Decrece	7842.8	10149.4	7144.6	Tráfico escaso (-)
6	4519	Decrece	7273.7	9217.2	6468.4	Tráfico escaso (-)
7	6616	Crece	5345.1	8623.1	5924.3	Normal (+)
8	6979	Crece	6235.1	6717.3	3972.9	Congestión (+)
9	5251	Decrece	6756.2	7545.4	4924.7	Normal (-)
10	6469	Crece	5702.3	8061.5	5450.8	Normal (+)
11	4739	Decrece	6239.3	6953.9	4450.7	Normal (-)
12	6779	Crece	5188.8	7441.1	5037.6	Normal (+)
13	7535	Crece	6302.1	6356.8	4020.9	Congestión (+)
14	5671	Decrece	7165.4	7431.1	5173.1	Normal (-)
15	8631	Crece	6119.5	8269.8	6061.0	Congestión (+)
16	7772	Decrece	7877.4	7270.3	4968.7	Congestión (-)
17	11239	Crece	7803.5	9062.0	6692.8	Congestión (+)
18	4926	Decrece	10208.6	9197.4	6409.7	Tráfico escaso (-)
19	7194	Crece	6510.5	11598.9	8818.4	Tráfico escaso (+)
20	7648	Crece	6988.9	7895.4	5125.6	Normal (+)
21	6809	Decrece	7450.2	8366.9	5610.8	Normal (-)
22	12001	Crece	7001.1	8948.4	5952.0	Congestión (+)
23	13443	Crece	10501.4	8479.2	5523.0	Congestión (+)
24	12350	Decrece	12560.3	11993.3	9009.4	Congestión (-)
25	13118	Crece	13177.9	14068.5	11052.0	Normal (+)
26	11534	Decrece	13363.2	14656.9	11698.9	Tráfico escaso (-)
27	13512	Crece	13418.8	14853.1	11873.3	Normal (+)
28	11630	Decrece	13435.5	14881.9	11955.7	Tráfico escaso (-)
29	13026	Crece	13440.5	14895.5	11975.5	Normal (+)
30	13015	Decrece	13442.0	14895.9	11985.1	Normal (-)
31	12856	Decrece	13442.4	14887.6	11996.4	Normal (-)

de ajuste K igual a 1 (Fig. 4). Con los umbrales definidos y el patrón previamente conocido, el motor de inferencia genera conclusiones sobre el estado de la red. Si la observación analizada excede el umbral superior o es menor al umbral inferior, se infieren anomalías de congestión o subutilización de la red, respectivamente, además de indicar si la tendencia es creciente (+) o decreciente (-). Las conclusiones obtenidas para cada observación se muestran en la última columna de la Tabla VII. Como puede verse, para las últimas ocho observaciones (horizonte temporal pronosticado), el motor de inferencia notifica una anomalía de congestión creciente para la observación 24, y dos de tráfico escaso y decreciente correspondiente a las observaciones 26 y 28. Es importante tener en cuenta que la variación del parámetro K afecta directamente al nivel de restricción bajo el que operan los umbrales adaptativos. Tal y como se muestra en Fig. 5, en los valores más bajos de K aumenta la distribución de síntomas reportados relacionados con la identificación de congestión en la red o bajadas representativas del volumen de datos promedio en el tráfico monitorizado; en concreto, cuando K se aproxima a 0, éstos suponen más de la mitad de los informes emitidos. Sin embargo, a medida que decrece el nivel de restricción, disminuye su frecuencia

de aparición, llegando a representar aproximadamente el 10% de los etiquetados realizados. Nótese que en la actualidad existen diferentes estrategias de calibrado y adaptación de variables de ajuste similares, a los diferentes problemas relacionados con adaptar sensores a escenarios de monitorización no estacionales, recopilándose en [28] algunas de las más relevantes. Pero dada su complejidad y a menudo, dependencia de los casos de uso, profundizar en ello está fuera del alcance de este artículo, estableciéndose de esta manera una interesante línea de trabajo futuro.

## V. CONCLUSIONES

El trabajo presentado propone un marco de análisis para la inferencia de conocimiento en redes 5G, y se enmarca en el desarrollo del proyecto de financiación Europea SELFNET. Su principal objetivo es generar conocimiento, a partir de la información monitorizada y agregada en etapas previas de procesamiento, con el cual es posible generar conclusiones, expresadas como síntomas, acerca del estado actual y futuro de la red. Las métricas agregadas y eventos provenientes de los niveles de infraestructura física y virtual sirven además para identificar o desencadenar alertas de seguridad sobre un dominio de red específico. Para la adquisición de conocimiento, el marco

de análisis distingue las etapas de integración de casos de uso, descubrimiento, reconocimiento de patrones, predicción, cálculo de umbrales adaptativos, inferencia de conocimiento y notificación. El conocimiento obtenido sobre el contexto de operación permite que la capa de gestión autónoma (inteligencia) mejore y optimice las tareas de diagnóstico y toma de decisiones relacionadas con la gestión de incidencias, facilitando así el despliegue automático de contramedidas proactivas y reactivas que busquen mitigar los problemas de red detectados.

La aproximación realizada ha sido probada sobre un caso de uso real, donde se ha procesado información procedente de capturas de tráfico monitorizadas en la red de la Universidad Complutense de Madrid. Con este fin se ha procedido a su instanciación, y el componente de predicción ha sido validado a partir del estándar funcional resultante de la competición M3, demostrándose su capacidad de ofrecer pronósticos precisos. En los resultados obtenidos es posible observar cómo desde un ajuste muy sencillo, es posible inferir conocimiento para adaptar la red y su configuración a la evolución del volumen de tráfico observado. Además, el camino trazado permite distinguir diferentes líneas de investigación orientadas a su mejora, como por ejemplo el diseño de estrategias de autoajuste adaptadas a los escenarios 5G, la definición de síntomas más precisos, o la identificación de métodos que mejoren su robustez frente a técnicas de evasión, como por ejemplo los métodos de imitación (*mimicry*). También debe considerarse la necesidad de ampliar la experimentación a casos de uso más complejos, que presenten una mayor similitud con circunstancias potencialmente observables en redes de 5G. Es importante destacar que a día de hoy no existen colecciones de muestras ni trazas de tráfico con estas características, por lo que habrá que esperar a que estas tecnologías alcancen un mayor nivel de madurez para ser capaces de evaluar con precisión tanto el trabajo descrito en este artículo, como el resto de iniciativas de investigación en 5G.

#### AGRADECIMIENTOS



This work was funded by the European Commission Horizon 2020 Programme under Grant Agreement number H2020-ICT-2014-2/ 671672 SELFNET (A Framework for Self-Organized Network Management in Virtualized and Software Defined Networks).

#### REFERENCIAS

- [1] J.C. Hourcade, Y. Neuvo, R. Saracco, W. Wahlster, R. Posch, "Future Internet 2020: Visions of an Industry Expert Group", Panel Report, May 2009.
- [2] P.K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, A. Benjebbour, "Design Considerations for a 5G Network Architecture", IEEE Communications Magazine, vol. 52, n. 11, pp. 65-75, November 2014.
- [3] L.I. Barona López, A.L. Valdivieso Caraguay, M.C. Sotelo Monge, L.J. García Villalba, "Key Technologies in the Context of Future Networks: Operational and Management Requirements", Future Internet, vol. 9, n. 1, pp. 1-15, October 2016.
- [4] "ITU-T P-1201: Parametric non-intrusive assessment of audiovisual media streaming quality". <http://handle.itu.int/11.1002/1000/11727> (Accessed 14 July 2017).
- [5] A. Belmonte Martin, L. Marinos, E. Rekleitis, G. Spanoudakis, N. Petroulakis, "Threat Landscape and Good Practice Guide for Software Defined Networks/5G", January 2016.
- [6] D. Kreutz, F.M. Ramos, P.E. Verissimo, C.E. Rothenberg, S. Azodolmolky, S. Uhlig, "Software-defined Networking: A Comprehensive Survey", in Proc. of the IEEE, vol. 103, n. 1, pp. 14-76, December 2015.
- [7] S. Abdelwahab, B. Hamdaoui, M. Guizani, T. Znati, "Network Function Virtualization in 5G", IEEE Communications Magazine, vol. 54, n. 4, pp. 84-91, April 2016.
- [8] Q. Zhang, L. Cheng, R. Boutaba, "Cloud Computing: state-of-the-art and Research Challenges", Journal of Internet Services and Applications, vol. 1, n. 1, pp. 7-18, May 2014.
- [9] A. Imran, A. Zoha, A. Abu-Dayya, "Challenges in 5G: How to Empower SON with Big Data for Enabling 5G", IEEE Network, vol. 28, n. 6, pp. 27-33, December 2014.
- [10] N. Baldo, L. Giupponi, J. Mangues-Bafalluy, "Big Data Empowered Self Organized Networks", in Proc. of the 20th European Wireless Conference, pp. 1-8, Barcelona, Spain, May 2014.
- [11] "5G Americas", <http://www.5gamericas.org/es/> (Accessed 01 May 2017).
- [12] "IMT-2020 (5G) Promotion Group", <http://www.imt-2020.cn/en/introduction> (Accessed 01 May 2017).
- [13] W. Mohr, "The 5G Infrastructure Public-Private Partnership", ITU GSC-19 Meeting, 2015.
- [14] "Self-Organized Network Management in Virtualized and Software Defined Networks (SELFNET)", Project reference: H2020-ICT-2014-2/671672, Funded under: H2020. <http://www.selfnet-5g.eu> (Accessed 01 May 2017).
- [15] P. Neves, R. Calé, M.R. Costa, C. Parada, B. Parreira, J. Alcaraz-Calero, Q. Wang, J. Nightingale, E. Chirivella-Perez, W. Jiang, "The SELFNET Approach for Autonomic Management in an NFV/SDN Networking Paradigm", International Journal of Distributed Sensor Networks, pp. 1-17, December 2015.
- [16] M.R. Endsley, "Design and Evaluation for Situation Awareness Enhancement", in Proc. of 32nd Annual Meeting on Human Factors Society, vol. 32, no. 1, pp. 97-101, October 1988.
- [17] "CROWD Project (2013) Connectivity Management for EneRgy Optimised Wireless Dense Networks", Funded under: FP7-ICT. Project Reference: 318115, <http://www.ict-crowd.eu/> (Accessed 01 May 2017).
- [18] "T-NOVA Project (2013) Network Functions as-a-Service over Virtualised Infrastructures", Funded under: FP7-ICT. Project Reference: 619520. <http://www.t-nova.eu/> (Accessed 01 May 2017).
- [19] "UNIFY Project (2013) Unifying Cloud and Carrier Networks". Funded under: FP7-ICT. Project Reference: 619609. <http://www.fp7-unify.eu/> (Accessed 01 May 2017).
- [20] L. Xu, H. Assem, I.G.B. Yahia, T.S. Buda, A. Martin, D. Gallico, M. Biancani, A. Pastor, P. Aranda, M. Smirnov, "CogNet: A Network Management Architecture Featuring Cognitive Capabilities", in Proc. of the European Conference on Networks and Communications, pp. 325-329, June 2016.
- [21] "CHARISMA Project (2014) Converged Heterogeneous Advanced 5G Cloud-RAN Architecture for Intelligent and Secure Media Access". Funded under: H2020-ICT-2014-2. Project Reference: 671704. <http://www.charisma5g.eu/> (Accessed 01 May 2017).
- [22] "5G-Ensure Project (2014) Enablers for Network and System Security and Resilience". Funded under: H2020-ICT-2014-2. Project Reference: 671562. <http://www.5gensure.eu/> (Accessed 01 May 2017).
- [23] S. Makridakis, M. Hibon, "The M3-Competition: results, conclusions and implications", International Journal of Forecasting, vol. 16, n. 4, pp. 451-176, December 2000.
- [24] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, "Forecasting: Methods and Applications", John Wiley & Sons, 1998.
- [25] C. Forgy, "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem", Artificial Intelligence, vol. 19, n. 1, pp. 17-37, September 1982.
- [26] Drools, <http://https://www.drools.org> (Accessed 01 May 2017).
- [27] L.J. García Villalba, A.L. Sandoval Orozco, J. Maestre Vidal, "Advanced Payload Analyzer Preprocessor", Future Generation Computer Systems, doi.org/10.1016/j.future.2016.10.032, November 2016.
- [28] G. Ditzler, M. Roveri, C. Alippi, R. Polikar, "Learning in Nonstationary Environments: A Survey" IEEE Computational Intelligence Magazine, vol. 10, no. 4, pp. 12-25, november 2015.