

Received July 19, 2021, accepted September 19, 2021, date of publication September 22, 2021, date of current version September 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3114593

A 5G-Based eHealth Monitoring and Emergency Response System: Experience and Lessons Learned

KIRIL ANTEVSKI¹, LUIGI GIRLETTI¹, CARLOS J. BERNARDOS¹,
ANTONIO DE LA OLIVA¹, (Member, IEEE), JORGE BARANDA², (Senior Member, IEEE),
AND JOSEP MANGUES-BAFALLUY²

¹Telematic Engineering Department, University Carlos III of Madrid, 28911 Leganés, Spain

²Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), 08860 Castelldefels, Spain

Corresponding author: Kiril Antevski (kiril.antevski@uc3m.es)

This work was supported in part by the EU H2020 5GROWTH Project under Grant 856709, in part by Madrid Government (Comunidad de Madrid-Spain) through the Multiannual Agreement with University Carlos III of Madrid UC3M in the line of Excellence of University Professors under Grant EPUC3M21, and in part by the Regional Program of Research and Technological Innovation (V PRICIT).

ABSTRACT 5G is being deployed in major cities across the globe. Although the benefits brought by the new 5G air interface will be numerous, 5G is more than just an evolution of radio technology. New concepts, such as the application of network softwarization and programmability paradigms to the overall network design, the reduced latency promised by edge computing, or the concept of network slicing – just to cite some of them – will open the door to new vertical-specific services, even capable of saving more lives. This article discusses the implementation and validation of an eHealth service specially tailored for the Emergency Services of the Madrid Municipality. This new vertical application makes use of the novel characteristics of 5G, enabling dynamic instantiation of services at the edge, a federation of domains and execution of real on-the-field augmented reality. The article provides an explanation of the design of the use case and its real-life implementation and demonstration in collaboration with the Madrid emergency response team. The major outcome of this work is a real-life proof-of-concept of this system, which can reduce the time required to respond to an emergency in minutes and perform more efficient triage, increasing the chances of saving lives.

INDEX TERMS eHealth, 5G, vertical services, AR/VR, virtualization, orchestration, multi-domain.

I. INTRODUCTION

The new generation of mobile communications, 5G, is expected to bring significant improvements on many fronts: enhanced mobile broadband experiences to the end-user, ultra-reliable extremely low latencies to enable industry automation, autonomous driving, and massive machine-type communications, which will make the wireless Internet of Things a reality. But in addition to these highlight-worthy well-known use cases, there are many application areas that could benefit from 5G and associated technologies. One key example, with a clear and direct impact on society at large, is emergency services and healthcare.

Nowadays, emergency services depend on human intervention. A witness in the vicinity of the emergency will,

The associate editor coordinating the review of this manuscript and approving it for publication was Ravinesh C. Deo¹.

luckily, start the emergency procedure described as follows: (i) the witness calls the emergency number (112 in Spain) and explains the situation and the location (this explanation is subjective and prone to errors based on the background of the witness, since there is no available data of the patient's condition, also the location is subjective and very often referred to geographical items difficult to locate, e.g., next to the bakery), (ii) the operator at the 112 call center assesses the situation and decides which is the most suitable emergency response team, and, (iii) the emergency team is deployed.

In the city of Madrid, this procedure takes around 4 minutes, while the time an ambulance takes to reach the location is estimated to be around 8 minutes (depending on distance and traffic).¹ By analyzing the data provided by the

¹https://www.madrid.es/UnidadesDescentralizadas/Emergencias/Samur-PCivil/Samur/ApartadosSecciones/11_NoticiasNovedades/Ficheros/Jornada5G_SalvarVidas.pdf

Emergency services of Madrid, we realized that considerable improvements can be achieved in improving the efficiency of the emergency service by employing automatized detection systems.

The automatic detection of emergencies is an incipient business that will increase in the next years thanks to the new capacities in terms of massive connectivity of devices with low power consumption brought by 5G. The increasing trend on the use of smart wearable devices, together with the great advance in terms of portable medical monitoring and sensing, can be used to enable continuous monitoring of health parameters (e.g., heart rate, sugar blood level, blood pressure), and thus detect, and even predict, potential health issues in a personalized way.

Additionally, the introduction of 5G brings more opportunities to improve the quality of care of the emergency team. New Augmented Reality (AR) technologies allow for better treatments on-site, as well as to enable remote support from other medical teams, which may reduce the cost of the service (reduced emergency teams supported remotely). AR requires significant computing resources that cannot be pushed to the cloud, due to strict latency requirements. However, 5G capabilities in terms of low latency, rapid edge deployment, and high reliability can be used to make possible the use of AR services when treating an emergency situation. The 5G network also enables providing such service globally through service federation. In this way, Emergency Systems that are customers of a given 5G service provider can track and respond to emergencies of their patients everywhere by using resources of other providers.

With the goal of realizing an improved and automatized emergency service, this article proposes a design and realization of a 5G personalized health emergency system. It describes a real-life experience of the deployment of a system capable of detecting and responding to emergency situations in an automatic and personalized way while enriching the tools at hand of the emergency team by enabling the use of AR services at the location of the emergency, thanks to the dynamic network reconfiguration. In short, the system is capable of patients' live-monitoring so that when an accident occurs (e.g., irregular heartbeat, which is a sign of a possible cardiac arrest), the system triggers an alarm to send an emergency team to the emergency location while the network is reconfigured with the deployment of a new network service in order to support the emergency team with AR services on-site. The presented scenario has been completely implemented and it has been validated by the emergency services of the Madrid Municipality. As further validated, the time needed to detect and process the emergency, select the best team and send it to the right location can be mostly eliminated by enabling an automatic and personalized emergency detection system, which also removes the need for the witness. This reduction translates into an increase in the patient's chances of surviving.

The rest of the paper is organized as follows. Section II introduces related work about the role of 5G, AR, and edge

computing in eHealth. Section III describes the scenario tackled, before presenting the developed eHealth system (in section IV). Validation results, using a real prototype with the ambulance and firefighter services of Madrid, are presented in section V. Finally, section VI presents the main takeaways from this full 5G validation experience, while section VII concludes this article.

II. RELATED WORK

5G will allow new kinds of health care services that are not feasible with the current capabilities of network technologies. In [1], the authors summarized the application of health care enabled by the capabilities of new mobile network technology. They divided them into mainly four categories, namely online consultation, online health monitoring, remote diagnosis, and mobile robot surgery (part of this work falls under the online health monitoring category). One of the challenges identified for this category was the high density of devices required to monitor a large part of the population. 5G is able to cope with the increasing number of chronic patients being monitored, as it introduces higher connection density, higher bandwidth, and lower latency with respect to the previous network generation (i.e., 4G). In [2], the authors demonstrate the need for a 5G network (compared to the 4G network) to guarantee high efficiency and fast responses in a health monitoring scenario.

Different technologies combined with 5G networks aim to improve health services. The works in [3], [4] focus on applying deep learning and artificial intelligence (AI) to enhance the performance in heterogeneous networks, also tackling how to enhance health services in [5]. Other works [6], [7] offer solutions to improve the security in healthcare systems, while some of them focus on multiple heterogeneous networks settings [8].

Edge computing is a key part of the 5G concept. The possibility of placing computational resources closer to the user contributes to providing the low latency required by health care applications while opening the door to new patient-centric applications. Both [9] and [10] make use of the concept of edge to provide patient care. In [9], a remote patient monitoring system makes use of edge resources to reduce the bandwidth needs of the telemetry system used to monitor the patients. using a variety of sensors and cameras. Also, it shows how using the edge ensures the real-time constraints of webRTC. In [10], the authors present a framework to assess voice disorders through deep learning processing at the edge. In our work, we use the edge for two differentiated functions: *i*) deploying a network service implementing an AR supporting system, and *ii*) deploying a virtual local breakout point.

Recent researches show that AR has huge potential for applicability in the healthcare system, comprising user-environment interfaces, telemedicine, and education [11]. A key example is the possibility to show relevant patient health records on the head-mounted AR device without losing focus on the patient. In [12], authors developed

a smart AR application that supports healthcare professionals with procedure documentation and patient information during wound treatments. In addition, they evaluated the interest of their work among healthcare professionals, who showed to prefer AR-based documentation systems with respect to the current documentation procedures (i.e., books, tablets, or smartphones). However, they only considered scenarios in which the AR application runs locally on the AR device.

Reference [13] evaluates the use of a particular AR device to assess its performance in a disaster scenario. Similar to the previous case, the AR application runs locally in the AR device, which additionally degrades the battery lifetime. In the study [14], AR devices are used to triage patients in a disaster event. In the study, the AR device operates in two modes (i) algorithm-assisted and (ii) with telemedical support from a remote professional. The results showed 90% of accurate triage. The study emphasizes the (WiFi) connectivity limitations, especially in the telemedical support case, and the low battery durability of the AR devices.

III. THE SCENARIO: 5G PERSONALIZED HEALTH EMERGENCY SYSTEM

In this section, we identify several areas where emergency services can be significantly improved by the use of new communication technologies enabled by next-generation mobile networks. On top of that, we present an eHealth emergency solution scenario, which we implemented as described in Section IV.

A. eHealth IMPROVEMENTS FOR SAVING LIVES

eHealth is defined as the delivery of health services by means of information and communication technologies (ICT). The goal is to improve the information flow between the actors involved (e.g., patients, paramedics, hospitals, doctors, surgeons), supported by ICT. Mobile health, which is a component of eHealth, is defined as a medical health practice supported by wireless devices, including wearable medical devices, patient monitoring devices, and personal assistants [15]. Adults are becoming more concerned and take measures for continuous health monitoring by investing in mobile health devices [16]. With the clear goal in mind of exploring how ICT can enhance the emergency response services, we have group different improvement opportunities as follows:

1) EMERGENCY RESPONSE TIME AND REAL-TIME DATA

In addition to preventing cardiac arrests through constant monitoring, reducing an emergency team's response time and being able to reach the exact patient's location clearly contributes to saving more lives. For example, the reduction of the emergency response time significantly lowers the door-to-balloon time.² According to [17], the guidelines

²The time period between the moment a patient with a possible acute heart-attack enters an Emergency Room and he/she undergoes balloon angioplasty surgery.

suggest a door-to-balloon time of fewer than 90 minutes. The study points out few effective strategies in reducing the door-to-balloon time. Some of them are (i) having a single call to a central operator, and (ii) providing real-time data feedback to emergency and catheterization laboratory prior to arrival. A similar study [18] analogously concludes that the major delays are due to reaching the patient and moving her to the closest hospital.

2) CONNECTIVITY, DURABILITY AND PERFORMANCE

As mentioned, technologies such as AR, have been proven to improve emergency services [14]. However, the main requirements to make AR useful in emergency scenarios are: (i) to have a stable and high-bandwidth wireless connectivity, and (ii) long battery duration. Various works [19], [20] suggest that computational offloading at the Edge can reduce energy consumption by 90%, while improving the overall performance of the devices. On top of that, Ultra-Low Latency Communication (URLLC) slices in 5G networks are envisioned to address the connectivity requirements for AR [21], [22].

B. SCENARIO DESIGN

Taking as a starting point the improvement considerations described before, plus the constraints imposed by the way emergency response teams operate and how cellular networks work, we arrived at the scenario shown in Figure 1. The ultimate goal, identified as critical by the Madrid emergency response team, is to develop a fully automatic and personalized emergency response system.

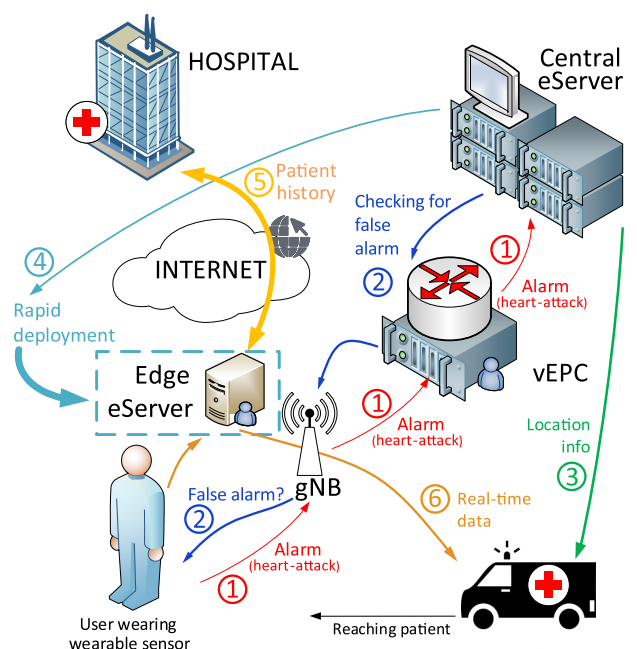


FIGURE 1. eHealth scenario.

To reach the goal, we set a simple scenario for which we designed the fully automatized system (explained in detail later). The scenario takes into account a continuously monitored patient for its vital signs (e.g., heart rate). Once the patient heart rate is abnormally low, an alarm is triggered which can be disregarded as a false alarm by the patient. If the patient does not mark the alarm as false within a short pre-configured amount of time, the system automatically dispatches an emergency team to the patient's location. In order to support the dispatched emergency team, the system deploys an AR system close to the patient's location.

To provide intensive health monitoring capabilities to an increasing part of the population, we rely on the 5G massive connectivity properties. A smart wearable (e.g., a smartwatch) is used to monitor the health of a person. This device is able to detect potential health issues, such as low blood sugar incidents or, as in our testing case, a heart-attack. Although 5G will support direct communication of these wearable devices to a central cloud through a low power communication, for this early stage of 5G deployment, we can assume the wearable is connected to a mobile phone application which periodically reports the health status and the patient's location to a central cloud (Central eServer, step 1 in Figure 1). At the functional level, there is not much difference between both solutions as far as the concepts presented in this article are concerned. If the monitored data reveals a potential (predicted) or actual health issue (e.g., heart rate down to zero), the Central eServer issues an alarm to the user mobile smartwatch or mobile phone (to check if it is a false alarm, step 2) while continuing the processing of the emergency event. This involves analyzing the health issue and the medical records of the person, deciding which might be the disease, and selecting the most appropriate team to deploy, considering both the time required to reach the location and skills that best address the emergency (e.g., a quick intervention medical vehicle, a regular ambulance, or the combined deployment of a firefighters team). The Central eServer automatically dispatches the selected emergency team (this can be canceled by the user at any time if the person notifies that it was a false alarm) to the location of the person (step 3). In this specific example, our system is able to deploy an AR service – for use of the emergency team – to improve the quality of care, by displaying geolocation and health information from the patient.

To provide high-performance, stable and durable AR service, the system requests the deployment of an emergency Edge eServer closer to the emergency location (step 4). This Edge eServer hosts the AR service helping the emergency team once deployed at the emergency location and may include patient health data used by the emergency team (step 5). This edge service is automatically deployed in a matter of a few minutes (with current technologies), while the emergency team reaches the indicated location. The deployed edge application establishes a connection to the emergency team and guides towards the location of the user by streaming an AR-marked pathway to the doctor's AR headset (step 6).

The edge application also obtains the user's health records and live-streams them on the doctor's AR headset together with real-time sensor data from the user's wearable. The AR headset is also used to live stream video to a remote medical team that can provide specialized support (if needed). Thanks to this, the paramedics' team can significantly increase their efficiency (e.g., faster triage and provide real-time feedback to the hospital), thus lowering the door-to-balloon time and increasing the probability of saving people's lives.

The use of AR technology in emergency scenarios has been already proposed, but it is now thanks to 5G that it becomes feasible to actually consider its wide use by emergency response teams since AR requires very low latency between the AR device and the AR server. As it will be proved later in this article, previous mobile networks (i.e., 4G) do not provide a latency low enough to guarantee good AR experiences. Our 5G-based solution is capable of dynamically instantiating an Edge eServer close to the location of the emergency and adapting the mobile network infrastructure to provide a low latency path to the newly deployed Edge eServer. In order to achieve this, a network service federation from different operators might be needed to satisfy the requirements of the AR service dedicated to the emergency case.

IV. THE SOLUTION: 5G-ENABLED PERSONALIZED HEALTH EMERGENCY SERVICE

This section describes the technical solution enabling the scenario described before. First, we provide an overview of the used 5G vertical service orchestration platform. Then, a detailed description of the developed emergency and AR applications is provided.

A. ORCHESTRATING NETWORK SERVICES IN 5G NETWORKS: 5G-TRANSFORMER

The H2020 5G-TRANSFORMER (5GT) project³ developed a platform with dynamic and flexible management features to automate the instantiation of multiple and heterogeneous services while satisfying the requirements coming from different vertical industries. These services can be concurrently instantiated over a shared infrastructure that includes multiple heterogeneous types of resources in terms of computing, storage, and networking, even spanning over multiple administrative domains. The 5GT architecture is made of four main building blocks, namely vertical slicer (5GT-VS), service orchestrator (5GT-SO), mobile transport and computing platform (5GT-MTP), and monitoring platform (5GT-MON). This architecture has been taken as a baseline by the H2020 5Growth project,⁴ which is extending it with additional features and innovations.

The 5GT-VS is the entry point for vertical industries to support the creation and management of network slices. The 5GT-SO oversees the end to end orchestration and the

³<https://www.5g-transformer.eu/>

⁴<https://www.5growth.eu/>

lifecycle management of NFV-network services (NFV-NSs) based on the available resources (compute, storage, and network) advertised by the underlying 5GT-MTP, which is the unified controller of transport stratum for integrated fronthaul and backhaul networks and computing and storage resources residing in multiple NFVI-Point of Presence. The 5GT-MON provides metrics to the 5GT platform so it can react and take decisions to adapt to network conditions and comply with service level agreements embedded in the NFV-NS request. A detailed description of the 5GT architecture can be found in [23].

As mentioned before, the 5GT architecture allows the deployment of network services spanning multiple administrative domains [24], known as network service federation (NSF). This is possible thanks to the capabilities of the 5GT-SO to orchestrate composite NFV-NSs (composed of multiple nested NFV-NSs). The NSF feature is essential for the deployment of the 5G personalized ehealth emergency system when and where needed, as shown in [25]. Let us just consider a simple example: the emergency services of a city municipality have a contract with a 5G operator to provide the patient’s monitoring and edge emergency NFV-NSs and the communication services used by all the emergency teams. This operator deploys with the 5GT platform most of its core network components and the monitoring NFV-NS (Monitoring-NS in Figure 2) in a remote cloud location because of operational reasons, e.g., not demanding latency constraints (AD1 in Figure 2). In case of an emergency, the Central eServer requests to the operator (by means of a query to the 5GT-VS) the instantiation of an edge emergency service (Edge-NS in Figure 2) connected to the monitoring NFV-NS close enough to the emergency location. A placement algorithm (in the 5GT-SO) [26]–[28] is in charge of deciding the placement of the Edge eServer based on (i) location constraints, (ii) information regarding the availability of local computation resources, and (iii) latency constraints of the AR application. The placement algorithm only computes

the ideal Edge eServer location over the operator’s local resources. However, there might be situations in which the operator does not have the infrastructure available in the proximity of the emergency location, requiring the use of infrastructure from a different operator offering the edge emergency service (AD2 in Figure 2).

Hence, excluding the NSF feature, an operator to implement our proposed solution would (i) require dedicated infrastructure, (ii) incur additional costs, and (iii) suffer from long implementation times. For example, a non-NSF implementation over the Madrid Municipality [27] (area of ~8000 km²) would require a dedicated infrastructure to satisfy the stringent AR latency requirements. That implies additional deployment & maintenance costs as well as longer implementation time.

B. HEALTH MONITORING AND AR APPLICATIONS

In addition to the network services and their orchestration logic, we developed the three applications needed for the 5G-enabled personalized health emergency service: (i) the monitoring application providing the heart rate measurements and location of users, (ii) the server application processing the monitored information, deciding if an emergency team has to be deployed and selecting the best one considering different information (e.g., time to reach the emergency based on the location of available ambulances), and (iii) the AR service, required to compute and stream the information reproduced on the AR headset (guidance to the physical location of the person, collection and representation of the relevant medical information at each moment, and video streaming to a remote medical team to better excel the patient’s triage).

The monitoring application is based on a smartwatch streaming heart rate data continuously to a 5G smartphone via ANT+ (Bluetooth could also be used). The smartphone is connected through 5G Non-Standalone (NSA) to the 5GT system and continuously sends new data to the Central eServer (see Figure 1).

The Central eServer itself is continuously checking the state of the patient based on the information received, detecting (or even predicting) when an emergency occurs, and contacting the person (to detect potential false alarms). A false alarm occurrence is discussed in Section VI. The steps performed by the Central eServer while attending the emergency are: (i) contacting the closest ambulance using the legacy emergency location system of the Madrid municipality (based on a GPS Fleet Navigation API), and (ii) triggering the instantiation of the edge emergency network service providing networking and computational resources required to support the emergency team upon their arrival to the patient’s location.

The Edge eServer provides remote rendered AR/VR video flow streamed through a 5G smartphone to the AR headset carried by the emergency team (we use Microsoft Hololens v1). The reason for this setup is the current lack in the market of AR headsets with 5G modems. Once the ambulance arrives at the emergency location, the Edge eServer starts streaming guidance information to the Hololens,

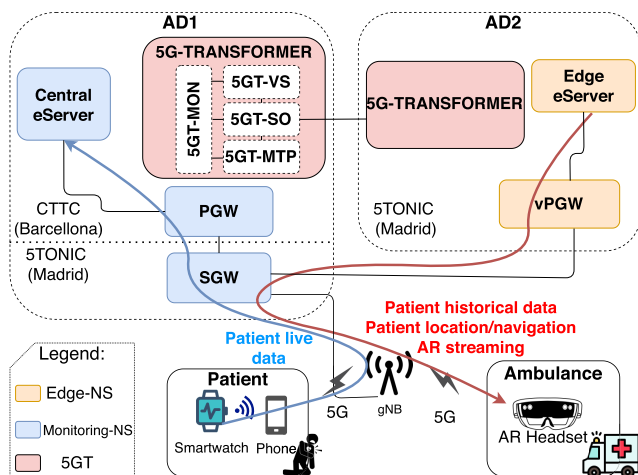


FIGURE 2. eHealth system deployment.

indicating directions to reach the patient location. When the team reaches the patient, medical information is displayed on the Hololens. This information is selected based on its temporal relevance and availability (e.g., results from historical blood tests) aimed at facilitating the decision flow of the medical team. This leads to more organized patient transportation along with a feedback video streamed from the Hololens, thus enabling real-time remote support from other remote medical teams or specialists in nearby hospitals.

The monitoring application is implemented using Android studio, using ANT+ API to gather information from the wearable and REST services to request functions from and to push data to the server. The Central eServer runs behind Apache HTTP server and it consists of a set of REST APIs, functions developed in PHP and GPS Navigator (Tomtom) APIs are used to find and contact the closest ambulance to the patient location.

The AR application is developed using Unity 2019 and built as a Universal Windows Platform application, it receives patient and paramedics position using the GPS location of the 5G smartphones. The streaming from the Edge eServer to the Hololens is implemented with the Holographic Remoting API provided by Windows Mixed Reality Toolkit (MRTK) [29]. The Hololens receives the stream using the MRTK native application, Holographic Remoting Player. The AR navigation was implemented using Mapbox [30]. Note that we did not implement any additional algorithms to assist the triage, as mentioned in [14]. The Hololens itself is able to capture uplink real-time video stream, while a hospital team is able to push feedback augmented information in the Hololens, assisting the emergency medic in real-time.

V. VALIDATION RESULTS

This section describes the experiments performed to assess the validity of our design and its usefulness for the emergency system of Madrid. To demonstrate the feasibility of the system, we deployed an end-to-end system as described in Figure 2, including a smartphone (Samsung S10) connected to a cellular 5G NSA network (provided by Ericsson BB630 baseband and Advance Antenna System AIR 6488) shown in Figure 5, the virtualized core network modules (implemented using the OpenEPC framework), a multi-domain 5G orchestration system (using the 5GT stack with one provider domain using Cloudify⁵ and the other one OpenSource MANO⁶ as coreMANO platforms) and the different applications required (both at the end-user and server sides). Validation was done by demonstrations/drills involving real emergency response teams with ambulances, medical staff and firefighters.

The location of the emergency is the Institute IMDEA Networks, host of the 5TONIC lab. The network functions of the different involved network services were deployed at 5TONIC and also at CTTC premises in Barcelona,

allowing us to resemble a scenario of a mobile operator providing services in Madrid, but having some of their core network entities in Barcelona (with a geographical distance of around 650 Kms). Following Figure 2, the RAN is deployed in 5TONIC, while the core network and eHealth Central eServer are deployed at CTTC. This geographical distance accounts for 10-15ms measured one-way delay in the communications between a UE and the Central eServer, which we will prove to be critical in order to deploy AR services.

The first step in our experimentation was to evaluate the service deployment time upon an emergency occurrence. More specifically the time it takes for the deployment of the Edge eServer at 5TONIC premises using the NSF. The bar chart in Figure 3 sums up the average time of all phases included in the Edge eServer deployment: (i) VNFs deployment, (ii) Connectivity establishment, and (iii) Service integration. The summed average deployment time is 7 minutes for a set of 10 deployments that we performed in 5TONIC.

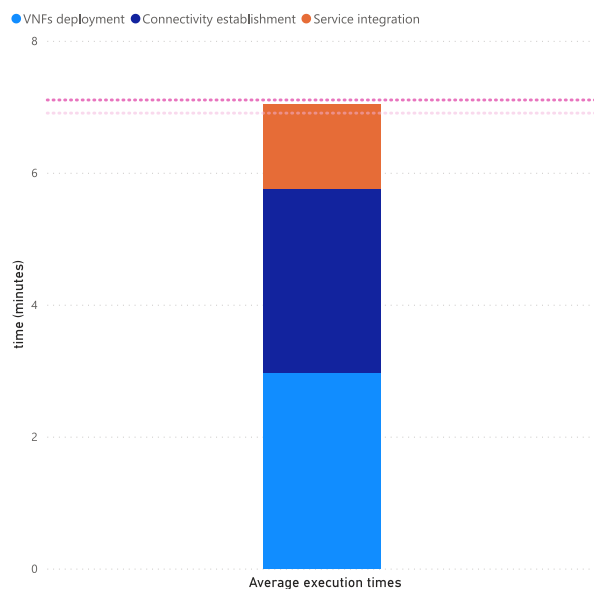


FIGURE 3. Service federation deployment time.

To further evaluate the feasibility and usefulness of our design, we analyzed the duration of every emergency operation that occurred in the Madrid Municipality from 01/01/2019 to 31/12/2019. The dataset has been exclusively provided by the Emergency Services of the Madrid Municipality, excluding any patients' private information. The graph in Figure 4 represents the cumulative distribution function (CDF) of the duration of every emergency. The duration time for each emergency is measured from the moment the emergency team/unit accepts the operation until it reaches the emergency location. Given the maximum deployment time of 7.1 minutes, there is a probability over 0.55 that the

⁵<https://cloudify.co/>

⁶<https://osm.etsi.org/>

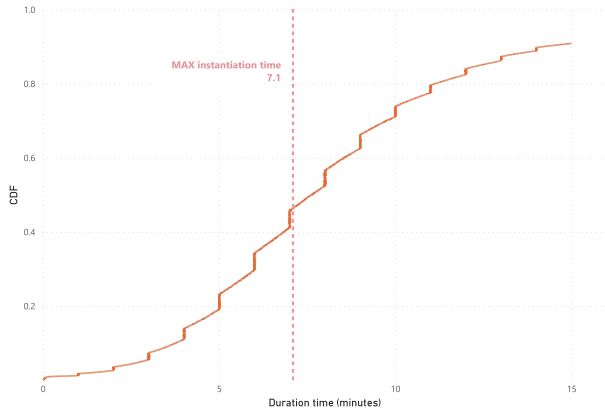


FIGURE 4. CDF of the duration from the moment that an emergency team accepts an emergency to the moment it reaches its location.



FIGURE 5. 5G NSA radio in the 5TONIC lab.

AR application is deployed and ready to be used by an emergency team upon arrival at an emergency site.

As mentioned before, the emergency average response time in Madrid is around 12 minutes, including around 4 minutes to issue the alert (receive the alarm and allocate the appropriate medical resources). The remaining portion is the time required to achieve the patient location shown with the CDF in Figure 4. In this context, the automatic detection of the emergency reduces the first 4 minutes almost to zero. Also, by the time the ambulance arrives at the emergency location, the Edge eServer will be up and running. This highly improves the response time which results in increasing the number of lives saved and reducing the number of side effects of a stroke.

The next step in our experiment was to evaluate the delay in the connection between the Hologens device and the server performing the AR computation, considering 4G and 5G technologies and the availability of local computing resources through federation. Note that if 5TONIC (local)

computing resources are available for federation, the AR service can be instantiated in the Edge eServer at the 5TONIC site. In other cases, the AR service cannot be deployed close to the emergency location, and therefore the AR minimal latency requirements are not met, e.g., if the AR service is deployed in CTTC central location. Table 1 summarizes the average one-way delay (OWD) measurements in the different configurations. From the obtained results, it is clear that in order to achieve an optimal AR service we need a 5G network connectivity with the AR application deployed using federated service close to the emergency location. In the case of 4G with local federated service and 5G without federated local service, the measured latency is too close to the minimal AR requirements.

TABLE 1. Average OWD of each scenario.

| Technology | Latency |
|------------------------------------|---------|
| 4G without federated local service | 30 ms |
| 4G with federated local service | 18 ms |
| 5G without federated local service | 13 ms |
| 5G with federated local service | 5 ms |
| Wifi | 2 ms |

For that reason, the final step in our experimentation has been to quantitatively assess the Quality of Experience (QoE) of the AR user, which is extremely sensitive to delay. To do so, we measured the average Frames Per Second (FPS) achieved at the Hologens on the different streaming settings. For example, X frames per second are streamed from the AR application to the Hologens, however, due to latency and packet loss, only $Y \leq X$ can be reproduced into the Hologens goggles. To capture the effects of latency on the FPS, we implemented an application module, using the diagnostic tool of MRTK, providing statistics about the actual frame rate sampled every half second. In this way, it was possible to analyze the average FPS achieved in the different streaming settings. Also, considering the caching, the frame prediction, and optimization of the Hologens, we performed our tests for a short time while moving into the AR world. In this way, it was possible to appreciate the real FPS experienced by the user. Obtained results are shown in Figure 6, where we denote the availability of local infrastructure for federated service with the label 5GT. It can be concluded that 5G is a clear must to have to achieve the best performance (being 60 FPS is the maximum frame-rate achievable by currently available AR devices).

Although the difference between 50 FPS and 60 FPS may seem small for the reader, it is important to highlight that FPSs are critical for AR applications. Any difference in the FPSs makes a huge difference in the experience of the user since head's movement tracking introduces a latency which is clearly visible for FPSs below 60.

Not only frame rate is affected by latency, but another problem that may occur is also object misplacement. Indeed, more than 10ms of delay leads to object misplacement of at least three degrees [31] in mobile AR. Figure 7 highlights

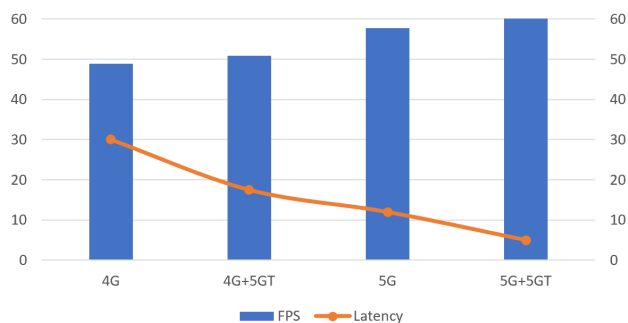


FIGURE 6. Average FPS achieved in the different scenarios.

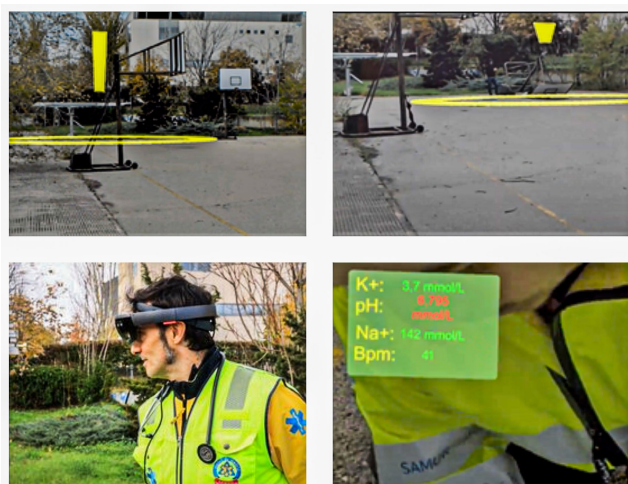


FIGURE 7. The object misplacement in the 4G without local edge scenario (top-left), a correct object placement in the 5G with local edge scenario (top-right), sanitary staff wearing the Hololens (bottom-left), patient health report shown on the Hololens (bottom-right).

the misplacement of 3D objects in the real world experience in the 4G scenario, compared to the correct position of the objects in the 5G one. The arrival point indicator is moved by various degrees from the original position in the test scenario using 4G without local edge (top-left), while in the best scenario, using 5G with a local edge, the indicator is placed correctly in the real world (top-right). Thus, enforcing the need for the proposed solution for mobile AR applications.

VI. LESSON LEARNED

This section lists the lessons learned during the implementation, integration, and deployment of the end-to-end eHealth system and network service development. We divided them by application-related and network-related.

A. APPLICATION-RELATED LESSONS

- *Do not always trust your phone's Geo-location.* The GPS location of smartphones has very high variability. In our work, an average of over ten samples of the coordinates was used to stabilize it. Hopefully, 5G will bring a more efficient localization mechanism.

- *Always synchronize the orientation.* The AR headset used in this work often lost its orientation when operating in dark environments. As consequence objects are misplaced in the real world by various degrees. There is a need for a fast and continuous synchronization mechanism of orientation tracking at the application level.
- *Choose carefully your GPU.* The remote rendering of the AR scene needs a server (Edge eServer) with powerful GPUs to stream the AR experience to the AR headset in real-time. In order to dynamically instantiate the Edge eServer, the GPU must be virtualized, a feature not supported by every GPU (including recent ones).
- *Need for AR feedback on AR stream reception.* We experienced that in some cases, when the latency is too high (over 100ms) and the bandwidth is too low (less than 8Mbps) the device does not get any AR input, without the server getting any notification/error. This needs to be improved to make the system more reliable.
- *False-alarms.* The occurrence of false alarms is a common and well-studied problem. However, there is no clear solution to approach it [32]. According to [33] almost 25% of the health emergency calls are false alarms, which avoiding them can produce immense savings. There are some models that can reduce these unnecessary calls [34], however, the authors warn that focusing on minimizing the false alarms can lead to an increase of more severe outcomes, even deaths. In our view, employing a limited timer to signal a false alarm (so once it expires, the emergency team is dispatched) could be sufficient to lower the number of false alarms without increasing the patients' risk.

B. NETWORK-RELATED LESSONS

- *Some VNFs require function-specific management and platform adaptations.* To exploit the capabilities of the 5GT platform, network services and VNFs deployed on top of the 5GT infrastructure might need to be adapted. Although this goes a bit against the virtualization principle of services and functions to be platform agnostic, as of today many virtualized functions (such as OpenEPC) have been designed with some platform assumptions in mind. In our tests, we had to deal with some specific VNF configurations in order to be able to instantiate and manage services over the 5GT platform.
- *VNFs are not just virtual machine images.* As an example, the EPC software used imposes a rigid IP and MAC addressing scheme, hindering its deployment in generic scenarios. This may prevent the use of this EPC stack in interoperable scenarios, where VNFs may be provided by different VNF vendors or pose difficulties to interact with physical equipment (PNFs), like the RAN component. Additionally, the design constraints of the associated VNFs required the introduction of ad-hoc operations to effectively enable the connection among the PNFs and VNFs, because some interactions could not be captured in the associated NSDs.

- *NSF helps in ubiquitous emergency handling.* As already mentioned in section IV-A, the NSF feature is an instrumental feature to enable the AR low latency requirements, and it provides an extended emergency) service geo-coverage while omitting the need of exclusive resources. This is proven through the validation results we obtained in section V.

VII. CONCLUSION

Nowadays, it is quite normal to find wearable devices capable of tracking sleep patterns, monitoring the heart rate, measuring the number of steps, or even perform an electrocardiogram. People with chronic diseases are taking these measurement devices to the next level, with patches able to measure glucose levels, connected insulin pumps, or wearable blood pressure meters. These devices will be complemented and augmented in 5G, one of its main characteristics being the focus on massive machine-type communications. Therefore, we expect all these devices to be connected to eHealth services, provided by public or private companies, which will perform continuous monitoring in order to detect anomalies and act upon them.

In this work, we have departed from this assumption and implemented a real use case showcasing the impact of 5G in the emergency service of Madrid. The deployed service aims at improving the quality of care in two ways: (i) reducing the time required to detect an emergency, while removing the variable of a human witness, (ii) providing location and health-related data to the emergency team for increased triage efficiency through augmented reality deployed in the field, and deliver real-time data back to hospitals to ameliorate surgery preparations.

These two services leverage the new capabilities of 5G, including not only the high-bandwidth and low-latency connectivity, but also the orchestration, federation, and dynamic instantiation of virtual functions at the edge of the network.

The use case was tested and validated by a real emergency team, showing that decreasing the response time by 30% is possible. Since every second is relevant when responding to emergencies, we believe the designed use case showcases and exemplifies the future evolution of emergency services.

ACKNOWLEDGMENT

The authors would like to thank SAMUR-PC for providing anonymized data improving the design and development of the work.

REFERENCES

- [1] H. Ullah, N. G. Nair, A. Moore, C. Nugent, P. Muschamp, and M. Cuevas, "5G communication: An overview of vehicle-to-everything, drones, and healthcare use-cases," *IEEE Access*, vol. 7, pp. 37251–37268, 2019.
- [2] J. Lloret, L. Parra, M. Taha, and J. Tomás, "An architecture and protocol for smart continuous eHealth monitoring using 5G," *Comput. Netw.*, vol. 129, pp. 340–351, Dec. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1389128617302189>
- [3] T. Mohammed, A. Albeshri, I. Katib, and R. Mehmood, "UbiPriSEQ—Deep reinforcement learning to manage privacy, security, energy, and QoS in 5G IoT HetNets," *Appl. Sci.*, vol. 10, no. 20, p. 7120, Oct. 2020.
- [4] N. Janbi, I. Katib, A. Albeshri, and R. Mehmood, "Distributed artificial intelligence-as-a-service (DAIaaS) for smarter IoE and 6G environments," *Sensors*, vol. 20, no. 20, p. 5796, Oct. 2020.
- [5] T. Muhammed, R. Mehmood, A. Albeshri, and I. Katib, "UbeHealth: A personalized ubiquitous cloud and edge-enabled networked healthcare system for smart cities," *IEEE Access*, vol. 6, pp. 32258–32285, 2018.
- [6] E. M. Abou-Nassar, A. M. Iliyasa, P. M. El-Kafrawy, O.-Y. Song, A. K. Bashir, and A. A. A. El-Latif, "DITrust chain: Towards blockchain-based trust models for sustainable healthcare IoT systems," *IEEE Access*, vol. 8, pp. 111223–111238, 2020.
- [7] A. A. A. El-Latif, B. Abd-El-Atty, W. Mazurczyk, C. Fung, and S. E. Venegas-Andraca, "Secure data encryption based on quantum walks for 5G Internet of Things scenario," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 1, pp. 118–131, Mar. 2020.
- [8] Y. Liu, J. Peng, J. Kang, A. M. Iliyasa, D. Niyato, and A. A. A. El-Latif, "A secure federated learning framework for 5G networks," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 24–31, Aug. 2020.
- [9] H. Moustafa, E. M. Schooler, G. Shen, and S. Kamath, "Remote monitoring and medical devices control in eHealth," in *Proc. IEEE 12th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2016, pp. 1–8.
- [10] G. Muhammad, M. F. Alhamid, M. Alsulaiman, and B. Gupta, "Edge computing with cloud for voice disorder assessment and treatment," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 60–65, Apr. 2018.
- [11] B. W. Munzer, M. M. Khan, B. Shipman, and P. Mahajan, "Augmented reality in emergency medicine: A scoping review," *J. Med. Internet Res.*, vol. 21, no. 4, Apr. 2019, Art. no. e12368. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/30994463>
- [12] K. Klinker, M. Wiesche, and H. Krcmar, "Digital transformation in health care: Augmented reality for hands-free service innovation," *Inf. Syst. Frontiers*, vol. 22, no. 6, pp. 1419–1431, Dec. 2020.
- [13] L. Carezzo, F. L. Barra, P. L. Ingrassia, D. Colombo, A. Costa, and F. Della Corte, "Disaster medicine through Google glass," *Eur. J. Emergency Med.*, vol. 22, no. 3, pp. 222–225, Jun. 2015.
- [14] A. Follmann, M. Ohlgs, N. Hochhausen, S. K. Beckers, R. Rossaint, and M. Czaplak, "Technical support by smart glasses during a mass casualty incident: A randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine," *J. Med. Internet Res.*, vol. 21, no. 1, Jan. 2019, Art. no. e11939.
- [15] MarketWatch. *Health Market: Industry Size, Growth, Analysis and Forecast of 2023*. Accessed: Mar. 5, 2020. [Online]. Available: <https://www.marketwatch.com/press-release/mhealth-market-industry-size-growth-analysis-and-forecast-of-2023-2018-09-11>
- [16] C. Meyer. *Social Life of Health Information*. Accessed: Mar. 5, 2020. [Online]. Available: <https://tincture.io/social-life-of-health-information-a7a7aac337c>
- [17] E. H. Bradley, J. Herrin, Y. Wang, and B. A. Barto, "Strategies for reducing the door-to-balloon time in acute myocardial infarction," *New England J. Med.*, vol. 355, no. 22, pp. 2308–2320, 2006.
- [18] E. B. Wu, N. Arora, A. C. Eisenhauer, and F. S. Resnic, "An analysis of door-to-balloon time in a single center to determine causes of delay and possibilities for improvement," *Catheterization Cardiovascular Interventions*, vol. 71, no. 2, pp. 152–157, 2008.
- [19] B. Shi, J. Yang, Z. Huang, and P. Hui, "Offloading guidelines for augmented reality applications on wearable devices," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1271–1274.
- [20] J. Dolezal, Z. Becvar, and T. Zeman, "Performance evaluation of computation offloading from mobile device to the edge of mobile network," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Oct. 2016, pp. 1–7.
- [21] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile internet," *IEEE Multimedia Mag.*, vol. 26, no. 1, pp. 21–30, Jan./Mar. 2019.
- [22] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G," *Ad Hoc Netw.*, vol. 4, pp. 169–177, Oct. 2018.
- [23] A. de la Oliva, X. Li, X. Costa-Perez, C. J. Bernardos, P. Bertin, P. Iovanna, T. Deiss, J. Mangues, A. Mourad, C. Casetti, J. E. Gonzalez, and A. Azcorra, "5G-TRANSFORMER: Slicing and orchestrating transport networks for industry verticals," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 78–84, Aug. 2018.
- [24] J. Baranda Hortiguela, J. Mangues-Bafalluy, R. Martinez, L. Vettori, K. Antevski, C. J. Bernardos, and X. Li, "Realizing the network service federation vision: Enabling automated multidomain orchestration of network services," *IEEE Veh. Technol. Mag.*, vol. 15, no. 2, pp. 48–57, Jun. 2020.

- [25] J. Baranda, J. Mangues-Bafalluy, L. Vettori, R. Martinez, K. Antevski, L. Girletti, C. J. Bernardos, K. Tomakh, D. Kucherenko, G. Landi, J. Brenes, X. Li, X. Costa-Perez, F. Ubaldi, G. Imbarlina, and M. Gharbaoui, "NFV service federation: Enabling multi-provider eHealth emergency services," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Jul. 2020, pp. 1322–1323.
- [26] K. Antevski, J. Martín-Pérez, and N. Molner, "Resource orchestration of 5G transport networks for vertical industries," in *Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Mar. 2018, pp. 158–163.
- [27] J. Martín-Pérez, L. Cominardi, C. J. Bernardos, A. de la Oliva, and A. Azcorra, "Modeling mobile edge computing deployments for low latency multimedia services," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 464–474, Jun. 2019.
- [28] J. Martín-Peréz, F. Malandrino, C.-F. Chiasserini, and C. J. Bernardos, "OKpi: All-KPI network slicing through efficient resource allocation," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 804–813.
- [29] *Mixed Reality Toolkit*. Accessed: Feb. 3, 2020. [Online]. Available: <https://github.com/microsoft/MixedRealityToolkit-Unity>
- [30] *Mapbox*. Accessed: Feb. 3, 2020. [Online]. Available: <https://www.mapbox.com>
- [31] W. Pasmán and F. W. Jansen, "Distributed low-latency rendering for mobile AR," in *Proc. IEEE ACM Int. Symp. Augmented Reality*, Oct. 2001, pp. 107–113.
- [32] M.-C. Chambrin, "Alarms in the intensive care unit: How can the number of false alarms be reduced?" *Crit. Care*, vol. 5, no. 4, pp. 1–5, 2001.
- [33] E. A. Blackstone, A. J. Buck, and S. Hakim, "The economics of emergency response," *Policy Sci.*, vol. 40, no. 4, pp. 313–334, Dec. 2007.
- [34] L. R. Barnes, E. C. Grunfest, M. H. Hayden, D. M. Schultz, and C. Benight, "False alarms and close calls: A conceptual model of warning accuracy," *Weather Forecasting*, vol. 22, no. 5, pp. 1140–1147, Oct. 2007.



KIRIL ANTEVSKI received the B.S. degree in telecommunication engineering from Saints Cyril and Methodius University of Skopje, Macedonia, in 2012, and the M.S. degree in telecommunication engineering from the Politecnico di Milano, Milan, Italy, in 2016. He is currently pursuing the Ph.D. degree in telematics engineering with University Carlos III Madrid (UC3M), Spain.

His research interests include the development of mechanisms to integrate and enhance NFV and MEC for 5G networks in dynamic and heterogeneous environments.



LUIGI GIRLETTI received the B.S. and M.S. degrees in computer science engineering from the University of Naples Federico II, Italy, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in telematics engineering with the University Carlos III of Madrid (UC3M), Spain.

His research interests include NFV and MEC solutions for 5G networks and artificial intelligence solutions to enhance networks performances.



CARLOS J. BERNARDOS received the degree in telecommunication engineering and the Ph.D. degree in telematics from the University Carlos III of Madrid (UC3M), in 2003 and 2006, respectively.

From 2003 to 2008, he was a Research and Teaching Assistant with UC3M, where he has been an Associate Professor, since 2008. He has published over 100 scientific papers in prestigious international journals and conferences.

Since 2005, he has been an Active Contributor to IETF, such as AUTOCONF, MEXT, NETEXT, DMM, MULTIMOB, SDNRG, and NFVRG working/research groups, being the coauthor of more than 30 contributions, ten RFCs, has co-chaired the IETF P2PSIP WG, and currently the Co-Chair of the IPWAVE WG and the Internet Area Directorate (INTDIR). He visited the Computer Laboratory, University of Cambridge, in 2004, and the University of Coimbra, in 2005. He has worked in several EU funded projects, being the Technical Manager of the FP7 MEDIEVAL and H2020 5G-Exchange Project and the Project Co-Ordinator of the 5G-TRANSFORMER and 5Growth Project. His current work focuses on virtualization applied to 5G networks.



ANTONIO DE LA OLIVA (Member, IEEE) received the degree in telecommunications engineering and the Ph.D. degree from the Universidad Carlos III Madrid (UC3M), Spain, in 2004 and 2008, respectively.

Since then, he has been an Associate Professor with UC3M. He is an Active Contributor of IEEE 802, where he has served as the Vice-Chair for IEEE 802.21b and the Technical Editor for IEEE 802.21d. He has also served as a Guest Editor of *IEEE Communications Magazine*. He has published more than 30 papers on different networking areas.



JORGE BARANDA (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Technical University of Catalonia, in 2008. He is currently a Senior Researcher with the Mobile Networks Department, CTTC, Barcelona. At CTTC, he has participated in several European, national, and industrial projects related with management and orchestration of SDN/NFV mobile networks, efficient routing strategies for mobile network backhauling, and novel wireless communication systems. He has coauthored over 40 different peer-reviewed journals and conference papers. His current research interests include management and orchestration of SDN/NFV mobile networks, wireless communications, wireless backhauling, network routing protocols, and network optimization.

His current research interests include management and orchestration of SDN/NFV mobile networks, wireless communications, wireless backhauling, network routing protocols, and network optimization.



JOSEP MANGUES-BAFALLUY received the bachelor's and Ph.D. degrees in telecommunications engineering from the Technical University of Catalunya (UPC), 1996 and 2003, respectively. He has been a Senior Researcher and the Head of the Communication Networks Division, Telecommunications Technological Center of Catalunya (CTTC), since 2013. He was also a Researcher and an Assistant Professor at UPC, from 1996 to 2003. He has over 20 years of experience in the networking field. He held various roles in several public funded and industrial research projects, such as EU 5Growth and 5G-REFINE (PI) on virtualization and automated network management, and also involving verticals, such as, automotive and eHealth. His research interests include mobile networks, machine learning for network optimization, and network functions virtualization. He was the Vice-Chair of IEEE WCNC 2018, Barcelona.

He held various roles in several public funded and industrial research projects, such as EU 5Growth and 5G-REFINE (PI) on virtualization and automated network management, and also involving verticals, such as, automotive and eHealth. His research interests include mobile networks, machine learning for network optimization, and network functions virtualization. He was the Vice-Chair of IEEE WCNC 2018, Barcelona.

...