Wireless Communications and Mobile Computing

Emergency Networks and Future Public Safety Systems

Lead Guest Editor: Maurizio Casoni
Guest Editors: Song Guo and Abderrahim Benslimane

Emergency Networks and Future Public Safety Systems

Wireless Communications and Mobile Computing

Emergency Networks and Future Public Safety Systems

Lead Guest Editor: Maurizio Casoni
Guest Editors: Song Guo and Abderrahim Benslimane

Editorial Board

Contents

Emergency Networks and Future Public Safety Systems
Maurizio Casoni, Song Guo, and Abderrahim Benslimane
Editorial (2 pages), Article ID 1647092, Volume 2019 (2019)

Tool for Recovering after Meteorological Events Using a Real-Time REM and IoT Management Platform
Yosvany Hervis Santana, David Plets, Rodney Martinez Alonso, Glauco A. Guillen Nieto, Nasiel Garcia Fernandez, Margot Deruyck, and Wout Joseph
Research Article (13 pages), Article ID 9767404, Volume 2019 (2019)

Sensing Coverage Algorithm of Sparse Mobile Sensor Node with Trade-Off between Packet Loss Rate and Transmission Delay
Kehua Zhao, Yourong Chen, Siyi Lu, Banteng Liu, Tiaojuan Ren, and Zhangquan Wang
Research Article (11 pages), Article ID 7080249, Volume 2019 (2019)

LMR and LTE for Public Safety in 700 MHz Spectrum
Aizaz U. Chaudhry and Roshdy H. M. Hafez
Review Article (17 pages), Article ID 7810546, Volume 2019 (2019)

Fast Multiattribute Network Selection Technique for Vertical Handover in Heterogeneous Emergency Communication Systems
Igor Bisio and Andrea Sciarrone
Research Article (17 pages), Article ID 8587932, Volume 2019 (2019)

Deploying a Reliable UAV-Aided Communication Service in Disaster Areas
Vicente Mayor, Rafael Estepa, Antonio Estepa, and German Madinabeitia
Research Article (20 pages), Article ID 7521513, Volume 2019 (2019)

Access Time Analysis of MCPTT Off-Network Mode over LTE
Yishen Sun, Wesley Garey, Richard Rouil, and Priam Varin
Research Article (19 pages), Article ID 2729370, Volume 2019 (2019)

Processing and Communication Delays in EWS: On the Performance of the Earthcloud Prototype
Martin Klapez, Carlo Augusto Grazia, Maurizio Casoni, Simone Zennaro, and Matteo Cozzani
Research Article (13 pages), Article ID 6206854, Volume 2019 (2019)

Next Generation Emergency Services Based on the Pan-European Mobile Emergency Application (PEMEA) Protocol: Leveraging Mobile Positioning and Context Information
Urban Sedlar, James Winterbottom, Bostjan Tavcar, Janez Sterle, Jaka Cijan, and Mojca Volk
Research Article (10 pages), Article ID 1408784, Volume 2019 (2019)

Resource Scheduling for Postdisaster Management in IoT Environment
J. Sathish Kumar and Mukesh A. Zaveri
Research Article (19 pages), Article ID 7802843, Volume 2019 (2019)

Multiobjective Based Resource Allocation and Scheduling for Postdisaster Management Using IoT
Meghavi Choksi and Mukesh A. Zaveri
Research Article (16 pages), Article ID 6185806, Volume 2019 (2019)

Textile Multiantenna Technology and Relaying Architectures for Emergency Networks
Estefanía Crespo-Bardera, Adrián Vega Delgado, Aarón Garrido Martín, Alfonso Fernández-Durán, and Matilde Sánchez-Fernández
Research Article (7 pages), Article ID 7292075, Volume 2019 (2019)

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 1647092, 2 pages
https://doi.org/10.1155/2019/1647092

Editorial

Emergency Networks and Future Public Safety Systems

Maurizio Casoni,[1] **Song Guo**,[2] **and Abderrahim Benslimane**[3]

[1]*University of Modena and Reggio Emilia, Modena, Italy*
[2]*The Hong Kong Polytechnic University, Hong Kong, China*
[3]*CERI/LIA University of Avignon, Avignon, France*

Correspondence should be addressed to Maurizio Casoni; maurizio.casoni@unimore.it

Received 29 July 2019; Accepted 29 July 2019; Published 14 August 2019

Natural or man-made disasters, CBRN (chemical, biological, radiological, and nuclear), can cause many casualties in urban areas and massive destruction in critical infrastructures. Terrorist attacks, especially in high-rise buildings, can be responsible for entrapment of a large number of people. Entrapment can also occur as a result of collapsed structures due to accidental or deliberate explosions (e.g., collapsed mines, technical failures, and confined spaces).

In past events, PPDR (public protection and disaster relief) agencies, e.g., fire brigades, ambulance service, and police, have always had many difficulties to effectively do their work because of technical and organizational issues. Interoperability among first responders belonging to different teams has always been very difficult, as well as coordination actions among the agencies [1]. In addition, managing and recovery of collapsed terrestrial telecommunication infrastructures has always been a major issue to solve in crisis events [2–4].

Ideally, future public safety systems should be interoperable, secure, and resilient for voice and data communications, supporting broadband communications and services. Therefore, the development and setup of advanced telecommunication and networking technologies for emergency networks as support of future public safety systems are among the most important tasks to face [5–7].

The purpose of this special issue is to publish original efforts in the ICT domain for effective future public safety systems.

Earthquakes have been responsible for more than 30% of the total fatalities from natural disasters in the last 30 years. A seismic alert system represents one of the most important measures to prevent and minimize earthquake damage. Klapez et al. present and evaluate the performance of Earthcloud, a cloud-based alert system helpful to reduce processing and communication delays.

Communications among first responders are fundamental for their work to provide help.

Crespo-Bardera et al. propose a two-hop relay network, merging MIMO textile technology at the first responders' jacket into the LTE-A cellular network. Their results prove the improvements in terms of network capacity and coverage of this novel architecture.

Future public safety systems have to operate over, possibly, dedicated radio resources.

Chaudry et al. present an overview of current LMR and emerging broadband LTE netwtorks to be deployed in the 700 MHz band. Their comparative study between current LMR networks and next LTE ones clearly shows the way for future public safety networks, even if several LTE services still do not meet mission-critical requirements, so that further studies and standardization activities are required.

However, public safety agencies around the world started migrating to LTE networks to support broadband communications. Sun et al. evaluate the performance of an off-network mission-critical push-to-talk device in direct mode communications over LTE. Their study may help to define configuration guidelines and end users to properly set the system.

Interoperability among agencies and countries is a mandatory requirement for future public safety systems. Sedlar at al. present the design, implementation, and results of a test bed for the 112 emergency service, which uses the Pan-European Mobile Emergency Application (PEMEA)

protocol. This test bed shows the feasibility of this approach, and it eases next developments of this system.

Emergency networks can employ wireless sensor networks, which can be sparse. Zhao et al. present a sensing coverage algorithm of sparse mobile sensor node, with tradeoff between packet loss rate and transmission delay.

After a severe meteorological event, telecommunication infrastructures and radio resources can be strongly damaged and made out-of-service. Therefore, it is very important to detect as soon as possible, possibly real time, the current availability of the radio spectrum. Santana et al. present a tool that senses and monitors in real time the radiofrequency spectrum and optimally places a number of sensors using Lora or other IoT technologies, to provide an emergency wireless coverage.

When telecommunication infrastructures are damaged by natural disasters, unmanned aerial vehicles (UAV) are also very useful to provide wireless coverage, given that they could be rapidly deployed. Mayor et al. investigate the optimal deployment of drones equipped with WiFi. Their proposal minimizes the number of drones required to provide reliable communication services.

A telecommunication infrastructure is composed of heterogeneous radio and mobile technologies. Therefore, in emergency scenarios, an efficient management of them is mandatory. Bisio et al. propose a decision maker in charge of performing network selection and handover decision. They evaluate several metrics through customized algorithms and show the best techniques for all the considered metrics.

In case of natural disasters, it is also very important to manage all available ICT resources at the best, to reduce delays and to improve information throughput among all actors involved in disaster recovery. Choksi et al. propose a solution for real-time resource allocation and scheduling to handle multiple objects in postdisaster situations. Their algorithm performs successfully in terms of maximal resource utilization and fulfilling the demand for different tasks at various places. Kumar et al. propose a resource scheduling algorithm for the postdisaster management in which they estimate the waiting time for the availability of resources using queueing theory.

With this special issue, we hope that readers will be interested in emergency networks and future public safety systems and they will find this issue helpful for their research and work.

Conflicts of Interest

Editors have no conflicts of interest to the assigned manuscripts when handling them and making decisions.

Maurizio Casoni
Song Guo
Abderrahim Benslimane

References

[1] H. Miller, R. Granato, J. Feuerstein, and L. Ruffino, "Toward interoperable first response," *IT Professional*, vol. 7, no. 1, pp. 13–20, 2005.

[2] S. Ghafoor, P. D. Sutton, C. J. Sreenan, and K. N. Brown, "Cognitive radio for disaster response networks: survey, potential, and challenges," *IEEE Wireless Communications*, vol. 21, no. 5, pp. 70–80, 2014.

[3] R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, "Public safety networks evolution towards broadband: sharing infrastructures and spectrum with commercial systems," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 24–30, 2016.

[4] M. Casoni, C. A. Grazia, M. Klapez, N. Patriciello, A. Amditis, and E. Sdongos, "Integration of satellite and LTE for disaster recovery," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 47–53, 2015.

[5] K. Gomez, S. Kandeepan, M. M. Vidal et al., "Aerial base stations with opportunistic links for next generation emergency communication," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 31–39, 2016.

[6] Y. Kyung, T. M. Nguyen, K. Hong, J. Park, and J. Park, "Software defined service migration through legacy service integration into 4G networks and future evolutions," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 108–114, 2015.

[7] M. Casoni, C. A. Grazia, and M. Klapez, "A software-defined 5G cellular network with links virtually pooled for public safety operators," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 3, p. e3092, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 9767404, 13 pages
https://doi.org/10.1155/2019/9767404

Research Article

Tool for Recovering after Meteorological Events Using a Real-Time REM and IoT Management Platform

Yosvany Hervis Santana,[1,2] **David Plets**,[1] **Rodney Martinez Alonso**,[1,2] **Glauco A. Guillen Nieto**,[2] **Nasiel Garcia Fernandez**,[2] **Margot Deruyck**,[1] **and Wout Joseph**[1]

[1]*Information Technology, Ghent University-IMEC, Ghent 9052, Belgium*
[2]*R&D Telecom, LACETEL, Havana 19200, Cuba*

Correspondence should be addressed to Yosvany Hervis Santana; yosvany.hervissantana@ugent.be

Received 28 March 2019; Revised 14 June 2019; Accepted 24 June 2019; Published 9 July 2019

Guest Editor: Maurizio Casoni

This paper is the design of a Radio Environment Map (REM) with a real-time tool to sense the radiofrequency spectrum and optimally places with Surrogate Modelling and Sequential Experimental Design tools a total of 72 SDR sensors in the selected area, using LoRa and/or NB-IoT technologies for networking. It permits the regulatory body to check the correct use of the assigned spectrum and constitutes a communication alternative in case of a catastrophic event, such as a hurricane or an earthquake, where radio and TV broadcasting play an important role in keeping people informed after such meteorological event. The radiobroadcast services use large antennas and high towers, making them vulnerable to such events. Regardless of the chosen technology, the IoT monitoring network will be more robust, since it uses small antennas and lower towers, and often a given area is covered by multiple base stations. The tool can be used to deploy new services in the nonserved area (e.g., 4G in the 700 MHz band at a lower cost or using TVWS techniques to provide communications and internet connection) and optimal interference management.

1. Introduction

Wireless communications and broadcasting play a vital role in connecting people around the world. These systems are used to inform, entertain, educate, and protect citizens. However, in developing countries, financial resources are not always available to have them operate efficiently.

Many developed countries introduce new broadband services, e.g., fourth-generation networks (4G), at higher frequencies (Long Term Evolution, LTE) band 38, 2570 MHz-2620 MHz). However, at higher frequencies, base stations will have a reduced coverage area, requiring much more base stations for the same service area. This densification of base stations is not feasible in underdeveloped countries because of the high cost. Moving these services to lower frequencies (e.g., 700 MHz) the numbers of base stations will be lower covering more area.

Some underdeveloped countries deploy their network in a progressive way based in coverage area without any planning tool. The major unsolved issue in efficiently sharing radio frequency spectrum between different services is related to the coverage area and radiated power by base stations.

Spectrum sensing performed by Cognitive Radio (CR) seems adequate due to its lower infrastructure requirement and wide application areas [1, 2]. In the case of CR, the spectrum-sensing task is to obtain the characteristics of spectrum usage. Moreover, it determines the type of signals that occupy the spectrum (digital or analog signals) and their features like the waveform, modulation, carrier frequency, and bandwidth, among others. However, this needs more precise signal analysis techniques which adds computational complexity to the detection algorithms [3].

CRs can avoid interference with primary users. Reliable sensing techniques identify the possible availability of spectrum opportunities to increase dynamic access to cognitive networks capacity. The most valuable parameters of the CR concept are the ability to sense, detect, learn, and be

aware of the radio operating environment, the state of the communication channel, the spectrum availability, the user requirements and applications, the local policies, the available networks (infrastructures), and the operating restrictions determined by the regulatory body. But the principal feature is its capacity of flexible autonomous autoreconfiguration [4].

Software Defined Radio (SDR) is one of the most significant and latest technologies already and profusely commercially in use for modern wireless communication-related systems. SDR, the hardware platform of almost all CR, can tune different frequency ranges and implement various (de-)modulation schemes and various standards in the same device/hardware by using a reconfigurable hardware and software system [5]. In SDR, a broader range of capabilities depends on elements which are basically software configurable [6]. The two major advantages of SDR are flexibility and easy adaptation. However, the deployments of SDR sensors technology with the capability of transmitting information are usually expensive [7]. Most recent advances on standardization and industrialization for the Internet of Things (IoT) technologies could contribute to solving the interconnection of the SDR devices.

Cuba is yearly hit by tropical storms (e.g., hurricane Irma in 2017) causing a lot of damage in the telecommunications infrastructure, destroying people's houses and making the circulation impossible. In this case, the recovery process should have good synchronization among the involved entities. Regardless of the chosen technology, the IoT monitoring network (using SDR sensors as sensing devices) will be more robust, since it uses small antennas and lower towers and often a given area is covered by multiple base stations. Having a Dynamic Monitoring Tool (DMT) able to detect in real time the radiofrequency spectrum and making suggestions with possible solutions is helpful to cover the damaged zone by the hurricane and keep the people informed.

The novelty of this paper is the design of a tool that senses and monitors in real time the radiofrequency spectrum and optimally places with Surrogate Modelling (SUMO) and Sequential Experimental Design (SED) tools a number of SDR sensors in the selected area, using LoRa and/or NB-IoT technologies for networking [1]. This tool will build a REM of the selected radiofrequency spectrum and use this information for the following purposes. First, it permits the regulatory body to check the correct use of the assigned spectrum. Second, it constitutes an alternative during recovery process in case of a catastrophic event, such as a hurricane or an earthquake, where radio and TV broadcasting play an important role in keeping people informed. The radiobroadcast services use large antennas and high towers, making them vulnerable to such events. Third, the tool will be useful for decision-making authorities and service providers, showing the real-time available frequencies and coverage interference about nonserved or underserved areas because of the damage caused by the event. Fourth, the tool can be used to deploy new services in the nonserved area (e.g., 4G in the 700 MHz band at a lower cost or using TVWS techniques to provide communications and internet connection).

This paper continues as follows. Section 2 presents the related work and the IoT platform solution. Section 3 proposes a network's architecture for our system. A real broadcasting scenario is considered, including the optimization for sensor position and IoT base stations. Section 4 shows the results for the network design, the feedback loop for different IoT platforms, and spectrum usage efficiency. Conclusions are presented in Section 5.

2. Related Work and IoT Platform Solutions

2.1. Related Work. A surrogate model is a cheap-to-evaluate replacement model of expensive, highly accurate computer simulations (e.g., sensors positioning). Hence, the SUrrogate MOdeling and Sequential Design tools are an alternative to deeply explore the design space by evaluating large amounts of samples. In [8], the authors use a surrogate model to simulate the behavior of RF circuits. In [9], the same authors use a surrogate model to find the best approximation of a LNA describing functions.

Using SED, all the data points are chosen at once and the modelling algorithm proceeds from there, without evaluating any additional samples later. In [10], the authors present a comparison and analysis of different space-filling sequential design methods, where the results are compared to traditional one-shot Latin hypercube designs. In [11], a comparison is made among different Sequential Experimental Design methods for global surrogate modeling on a real-world electronics problem.

Spectrum sensing, as a key enabling functionality in CRNs, needs to reliably detect weak Primary Radio (PR) signals of possibly unknown device types [4]. Spectrum sensing should also monitor the activation of primary users to vacate the occupied spectrum segments for secondary users. However, it is difficult for CR to capture such information instantaneously due to the absence of cooperation between the primary and secondary users. Recent research efforts on spectrum sensing have focused on the detection of ongoing primary transmissions by CR devices. Generally, spectrum-sensing techniques fall into two categories: wideband sensing and narrowband sense. Narrowband sensing splits coherent [12] and noncoherent detection. For coherent detection, no knowledge about the parameters of the primary signal is required, while noncoherent detection is the most appropriate one when the SDR has limited information on the primary signals (e.g., only the local noise power is known) [13]. Energy detector [14] is an example of noncoherent detection, while the waveform detector and the cyclostationary detector are coherent [3].

Collecting measurements for constructing the REM can be done by SDRs. Therefore, developing dedicated sensors to increase the quality of the REM must be considered. In [16], the effect of sensor geometries on Primary Users (PU) and the environmental parameter estimation are studied. Distributed spectrum sensing with Cognitive Radio Networks (CRNs) by exploiting sparsity is proposed in [17].

To obtain knowledge about the network, a generic approach to develop CR based on the REM is proposed in [18,

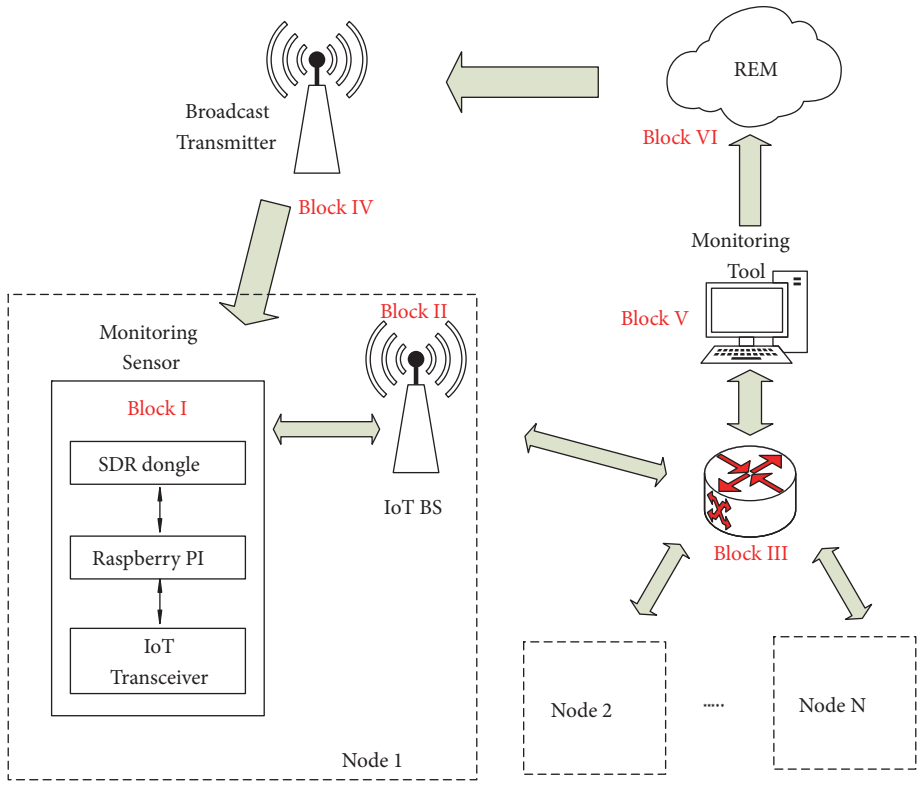


Figure 1: Dynamic network architecture.

19]. A REM is envisioned as an integrated database consisting of multidomain information, which supports global cross-layer optimization by enabling CR to "look" through various layers. The cognitive radio engine (CE) for various cognitive functionalities such as situation awareness, reasoning, learning, planning, and decision support can exploit the REM, as a vehicle of network support to CR. In [18], simulation results are presented. However, the authors do not consider different metrics for performance evaluation (e.g., the environmental noise interferences produced by cars, weather, and adjacent transmitters).

2.2. IoT Platform Solution. In recent years, IoT networks have increased quickly. The LoRaWAN (Long-Range Wide Area Network) is one of the most adopted IoT standards in the world [20]. Long-Range (LoRa) technology is generally implemented in the unlicensed 433 MHz and 868 MHz bands, with a channel bandwidth of 125 kHz. The physical layer (PHY) implements a Chirp Spread Spectrum modulation (CSS), which provides excellent robustness against interference [20].

Narrowband Internet of Things (NB-IoT) is a new cellular technology introduced in Third-Generation Partnership Project (3GPP) Release 13 for providing wide-area coverage for IoT. It includes relevant improvements for better performance of IoT applications. Narrowband-IoT (NB-IoT) allows flexibility by using a small portion of the traditional LTE network spectrum [21] requiring 180 kHz of bandwidth for both downlink and uplink. The choice of a minimum system bandwidth enables a number of deployment options (e.g., replacing one GSM carrier (200 kHz) with NB-IoT).

The air interface of NB-IoT is optimized to ensure harmonious coexistence with LTE, and thus such an "in-band" deployment of NB-IoT inside an LTE carrier will not compromise the performance of LTE or NB-IoT. An LTE operator also has the option of deploying NB-IoT in the guard-band of the LTE carrier by upgrading the software of the LTE BS.

3. Methods

3.1. Network Architecture. To implement a dynamic network, we designed a network architecture able to retrieve the required feedback data to the DMT in real time through an IoT feedback loop. Figure 1 shows a block diagram of the proposed network architecture.

The DMT (Block V in Figure 1) collects and analyses Quality of Service (QoS) (Block II, Figure 1) data retrieved from the sensing devices (Block I, Figure 1) to build the REM (Block VI, Figure 1) of the real propagation conditions in the covered area. The QoS data also allows detecting frequencies' interference from secondary services (e.g., adjacent transmitters) and available coverage and taking further actions (Block IV, Figure 1). The actions to take will depend on the scenario that the DMT is working on (e.g., normal conditions or after a disaster caused by a hurricane). In this way, a bridge between the broadcasting network and the DMT network (Block III, Node 1, Node 2, and Node N, Figure 1) was implemented. The

Table 1: Parameters to retrieve by the sensing device at the receiver location.

Parameters	Size	Unit
Channel	7	bit
Frequency offset	10	bit
SNR	7	bit
Bit Error Rate	13	bit
Total	*5*	*Bytes*

IoT network has to be designed and optimized to overlap with the broadcasting network coverage.

To build the sensing device, a Raspberry Pi (RPI), an SDR USB device, and an IoT transceiver were combined (Block I, Figure 1). The RPI has enough computational performance to drive both devices [22]. To detect Digital Terrestrial Multimedia Broadcast (DTMB) signals, the measurement device implemented by hardware using an SDR USB device must accomplish at least with the following parameters: (i) covering the radiofrequency UHF band (470-806MHz), (ii) bandwidth 6 MHz, and (iii) Signal-to-Noise Ratio (SNR) from 14 dB to 30 dB.

Table 1 lists the required parameters to implement the dynamic radiofrequency map and to provide feedback about interference issues to the DMT of Figure 1. The bit size for each parameter is software dependent. The frequency offset is required to correct the local frequency reference and achieve a higher measurement accuracy. A 5-byte packet is sent to the DMT. An IoT transceiver will collect the data packets from the sensing devices and later transmit it to the optimization servers every 5 minutes. Retrieved data is based on the 96-percentile for the previous 5 minutes. For LoRa packets of 10 bytes, the time on air with SF = 7 is around 40 ms [23]. Hence, the latency of the IoT network is not a critical constraint for this application.

3.2. Configuration and Scenario. First, define the proper link budgets for each technology, to sense the broadcasting network and to design and optimize the IoT network. Consider the IoT transceiver and the SDR sensor to be integrated as shown in Figure 1, Block I. Table 2 lists the most relevant link budget parameters for DTMB, SDR sensor, LoRa, and NB-IoT. A link budget accounts for all the gains, losses, and implementation margins in the transmitters, the receiver, and the propagation channel. Based on the link budget, it is possible to calculate the maximum allowable path loss (PLmax) for each technology in a certain scenario [24].

3.3. Broadcasting Network and Hurricane Effects. To make the dynamic map of the radiofrequency spectrum, a realistic suburban scenario in Havana, Cuba, and the currently deployed DTMB network (UHF band) was considered. Three DTMB transmitters (operating on 575 MHz, 677 MHz, and 689 MHz; the last two transmitters are situated in the same building) are covering an area of 50 km^2. Figure 2 shows the coverage area of the broadcast transmitters and available cellular base stations (BSs) (dotted line). To sense the

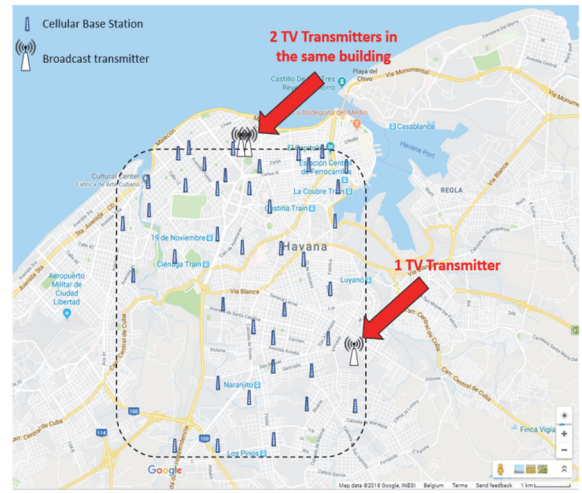


Figure 2: Broadcast transmitters, covered area, and cellular BSs (delimited with dotted line).

radiofrequency spectrum we deploy a number of sensors to cover the area and a number of IoT base stations (BSs) to collect the information from the sensors. The optimization and deployment of the sensors and IoT BSs will be described in Sections 3.5 and 3.6, respectively.

The OFDM parameters (including Frequency Sampling Factor) and bitrate of the broadcast transmitter were retrieved from the DTMB standard specifications [25]. Another specification of the broadcasting network, such as radiated power, radiation efficiency, frequency, bandwidth, antenna parameters, and receiver parameters, depends on the setup, network planning, and technology in use by the service provider (currently deployed network). Notice that the transmitter efficiency takes into consideration both the high-power amplifier and radiation system efficiency.

The shadowing standard deviation was retrieved from the Regulation for Digital Television Broadcast by the local regulatory authorities.

The wind speed in a tropical cyclone in the Caribbean could reach between 250 km/h and 400 km/h (e.g., hurricane Irma in 2017) [26]. The buildings have its structure prepared to support the wind speed, but the cyclone can destroy the windows, knock down the trees, and destroy the electrical service in the area making the circulation almost impossible. Let us suppose that, after the hurricane, two of the three transmitters and approximately 60% of the BSs are destroyed (Figure 3). The TV coverage zone and the operation of the DMT will be affected due to the lack of available infrastructure. The DMT will be helpful to check the spectrum availabilities. Hence, there are two options to get the DMT back: (i) giving priority to the BSs of the cellular network in the recovery process, deploying the DMT using an NB-IoT network, or (ii) deploying a LoRa network without dependencies on the existing or destroyed infrastructure. In both cases, the design of a network redundancy will be taken into account in case the base stations used for the DMT are destroyed by the hurricane.

Table 2: Link budget parameters.

Parameter	DTMB			LoRa	NB-IoT	Unit
Frequency	575	677	689	868	716	MHz
Radiated Power	50	60	58.45	14	23	dBm
Radiation efficiency	16.8	19.6	19.1	15.0	14.6	%
Bandwidth	6			0.125	0.015 0.180	MHz
OFDM Subcarriers	3780			-	12	-
OFDM Used Subcarriers	3744			-	12	-
Frequency Sampling Factor	0.420			0	1.536	-
Cell Interference Margin	0			0	2	dB
BS Antenna Height	40	145	145	20-31	30	m
Receiver Antenna Height	3			3	3	m
Receiver Antenna Gain	8			8	8	dB
Receiver Feeder Losses	0.6			0.6	0.6	dB
Noise Figure	3.5			6	3	dB
Shadowing Standard Deviation	7.5			7.5	7.5	dB
Receiver SNR	35			-20.0 -17.5 -15.0 -12.5 -10.0 -7.5	-12.6 3.0	dB
Bitrate	18274			0.24 0.44 0.97 1.75 3.12 5.46	0.02 204.8	kbps

3.4. IoT Infrastructure. To evaluate the required resources (i.e., infrastructure and spectrum usage) for the IoT feedback loop, we designed, optimized, and compared the two IoT networking solutions in the proposed scenario: LoRa and NB-IoT. Notice that the SigFox constraint of maximum packets delivered per day does not fit this application.

LoRa devices can radiate a signal level higher than 14 dBm (Class A end-devices), but due to the regulation of the maximum allowable radiated power in the 868 MHz, the maximum Equivalent Isotropically Radiated Power (EIRP) is limited to 14 dBm [27]. The maximum EIRP in NB-IoT end-devices is 23 dBm [28].

Using different configurations, LoRa BSs allow emulating up to 49 virtual channels [27]. Here, the maximum available physical channels only are considered (eight channels) [27]. The LoRa PHY layer implements a larger range of modulation schemes, allowing bit rates for a single channel from 0.25 kbps to 5.5 kbps [27]. The SNR is in the range from -7.5 dB to -20 dB. The spread spectrum modulation encodes each bit of information into multiple chirps. Hence, the spread spectrum processing gain allows receiving signal powers below the receiver noise floor.

For NB-IoT, we consider a joint deployment with LTE (band B8-900MHz) BS infrastructure, considering n-band mode. The occupied bandwidth per channel for LoRa is 125 kHz [27], and for NB-IoT is 180 kHz using a single LTE PRB [21]. The radiation efficiency of the NB-IoT power amplifier, radiation system, and the sampling factor will be the same as for LTE [29].

An additional 2 dB loss should be accounted for NB-IoT in the cell interference margin. This is because the LTE cell frequency distribution requires considering a permissible interference among nearby cells [24].

Table 3: The four different categories for sequential design and examples [15].

Input-based	Output-based	Model output-based	Model-based
Uses only input values from previous samples to determine next sample.	Uses input and output values from previous samples to determine next sample.	Uses previous samples and model evaluations to determine next sample	Uses previous samples and model properties and parameters to determine next sample.
Examples: (i) Random sampling (ii) Low-discrepancy sequences (iii) Sequentially nested Latin hypercubes (iv) Voronoi-based sampling (v) Monte Carlo Optimization-based sampling	Examples: (i) LOLA-Voronoi	Examples: (i) Adaptation to irregularities (ii) Slope, local optima and variance criteria (iii) Sequential Exploratory Experimental Design method (SEED) (iv) Model error sampling	Examples: (i) Kriging-based

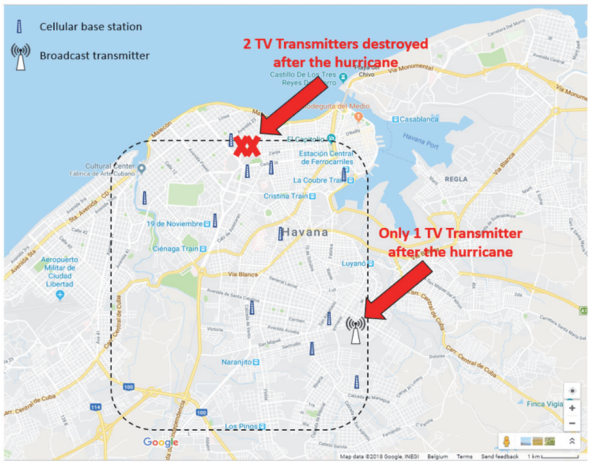


Figure 3: Broadcast transmitters, covered area, and cellular BSs (delimited with dotted line) after some meteorological event.

3.5. Sensing Network Optimizations. To optimize the sensing networks, several parameters need to be taken into account. Firstly, we have to cover an area of 50 km^2 in a suburban area with three TV transmitters working in the UHF band (470-806MHz). Hence, the propagation loss plays an important role at the moment to decide how many sensors we are going to use. To estimate its behavior, the Okumura-Hata model for propagation loss in a suburban scenario is considered, which is fully explained in [30].

Secondly, we have to calculate the required number of sensors to fit the area avoiding the errors trying to estimate the received power among adjacent sensors. To achieve that, the received power was estimated using the Okumura-Hata model in a suburban area varying the receiving distance in steps of 1 km. Figure 5 shows the results for the three transmitters in the area.

As can be seen, the received power drops dramatically in the first 10 km for all the transmitters. However, for the 575 MHz frequency and transmitted power of 50 dBm, after 5 km the received power drops under the TV sensitivity threshold established in a Cuban regulation for UHF band (-84 dBm) [31]. Hence, the proposition is to work with a resolution of one sensor per km^2.

To optimize the deployment the SUMO was used [32, 33]. The SUMO is a MATLAB tool that automatically builds accurate surrogate models (also known as meta-models or response surface models) of a given data source (simulation code, dataset, script, etc.) within the accuracy and time constraints set by the user. The tool minimizes the number of data points (which it chooses automatically) and tries to be as adaptive and autonomous as possible, requiring no user input besides some initial configuration [32, 33].

Usually, the simulations require expensive computational hardware. Hence, the SUMO tool is an alternative to deeply explore the design space by evaluating large amounts of samples. The goal of surrogate modeling is to find a model that mimics the original system's behavior but can be evaluated much faster. This function is constructed by calculating multiple samples at key points in the design space, analyzing the results, and selecting a model that approximates the samples and the system behavior.

Table 3 shows a wide variety of model types available; their limitations depend on the system that is being modeled. Popular choices are polynomial and rational functions [34], Kriging models [35], neural networks [36], and radial basis function (RBF) models [37]. These can be used to perform optimization and sensitivity analysis once the model is constructed [38].

Also, the SED [11, 39] was used trying to achieve a similar result and thus compare both deployments. The SED is a powerful tool for sequential Design of Experiments (DoE). In traditional experimental design, all the design points are selected up front, before performing any (computer or real-life) experiment, and no additional design points are selected afterward. This traditional approach is prone to oversampling and/or undersampling because it is often very

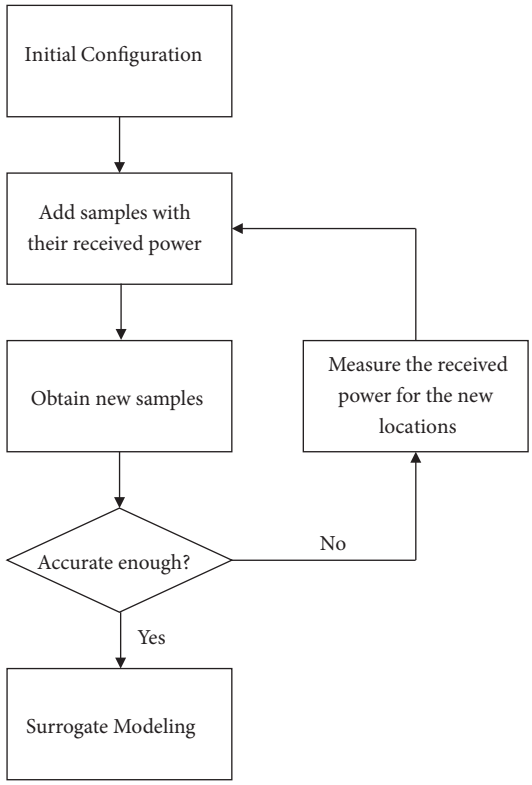


Figure 4: Flow diagram of the Sequential Design, including the SUrrogate MOdeling.

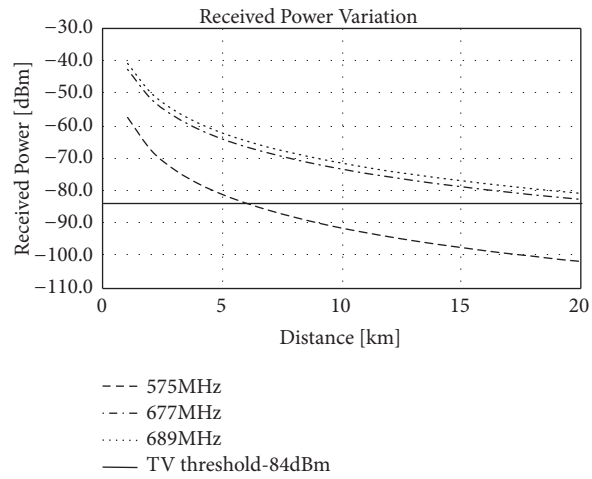


Figure 5: Received power variation using the Okumura-Hata model in steps of 1 km. The inputs parameters were taken from Table 2.

difficult to estimate the required number of design points in advance. The SED tool solves this problem by providing the user with state-of-the-art algorithms that generate an experimental design in a sequential way, i.e., one point at a time, without having to provide the total number of design points in advance. The SED was designed to be extremely fast and easy to use, yet very powerful [11, 39].

The traditional DoE is chosen based only on the available information in the first simulation, such as the input variables and the result of the measurements. This information is then added to the experimental design in the simulator, which evaluates all the information in the Surrogate Model. This is a one-shot run, where all the points are assessed at once, and the modeling algorithm proceeds without evaluating any additional samples [40].

Sequential Experimental Design (SED) improves on this approach by transforming the unique algorithm into an iterative process. SED methods analyze data (samples) and models from previous iterations to select new samples in areas that are more difficult to approximate, resulting in a more efficient distribution of samples compared to the traditional DoE.

Figure 4 shows the flow graph to work with both tools (SUMO and SED). In the first step (initial configuration), the coordinates (UMT format to SUMO tool and degree format to SED tool) of the area under evaluation need to be added. In addition, constraints are defined with, e.g., buildings, lakes, or rivers, not allowed areas to place the sensors. In both tools, these are known as "constraints." Finally, the number of samples to add in every iteration and the final number of samples to deploy need to be assigned. The rest of the process continues with a loop until finding the optimal number of sensors to deploy is defined. In the case of the SUMO tool, it starts adding known samples with their received power (dBm). If the number of samples does not satisfy the requirements, the tool advises new locations (samples), where the received power should be measured. There are then new samples and the whole process is repeated again. In the case of SED, putting the desired number of sensors to deploy is only needed and it will design the whole deployment.

Table 3 lists some of the available methods to work with SUMO and SED tools. A model is considered accurate enough when its root relative square error is lower than 0.05 [8]. The maximum number of samples allowed for each frequency was fixed at 65. In this paper, the output-based (LOLA-Voronoi) and model-based (Kriging-based) sampling were used because both use the input and output values from previous samples to determine the next samples. By adding samples after every simulation, the REM can be built as precisely as we want.

3.6. IoT Networks Optimizations. To optimize the IoT network, the power consumption of the dedicated feedback channel has to be minimized. To this aim, an IoT LoRa and an NB-IoT network (Table 2) are designed, optimized, and benchmarked.

To account for the minimally required infrastructure and optimize the network power consumption, the heuristic algorithm presented in [41, 42] and improved in [1] was used to reduce the number of base stations required in the IoT physical layer. Figure 6 shows the process for the design and optimization for minimal infrastructure and power consumption of the IoT feedback network. The design and the optimization are performed in two different steps (two heuristic cycles of the algorithm). In the original algorithm

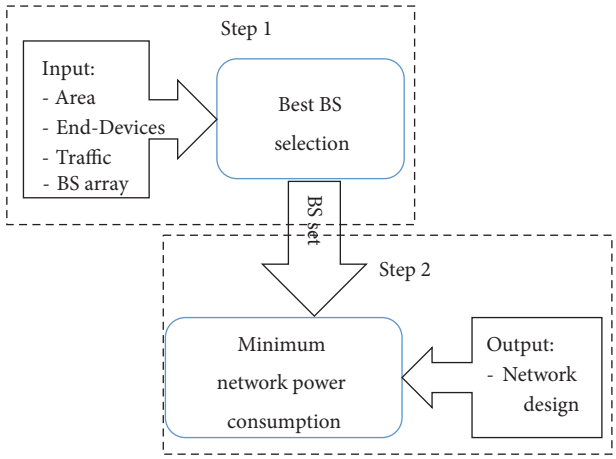


Figure 6: IoT network optimization process.

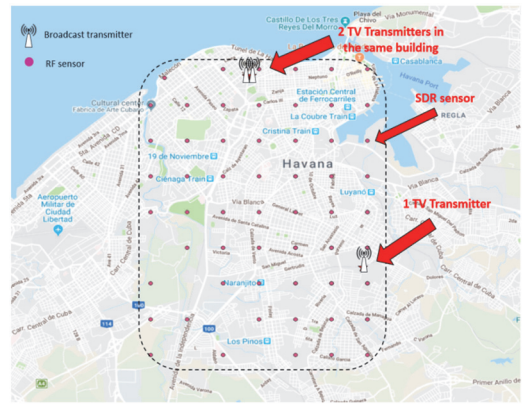


Figure 7: Sensing network deployment for the UHF band using SUMO and SED for optimal positioning and number of sensors for spectrum monitoring.

presented in [11, 41], only one cycle is present (no optimal BS location selection is implemented). The network design tool is capacity-based, meaning that the traffic density and end-devices density are input parameters. The software also receives as input parameters the target area and a number of possible BS geo-locations including the BS antenna height. Forty simulations are performed to assess the mean power consumption of the whole network, where the progressive average for all simulations is calculated to validate a proper estimation of the percentage of users covered.

In the first step (Figure 6), the best set of BSs among the whole set of possible BS locations is chosen. Forty-one available BS locations were considered in the scenario of Section 3.2. The software optimizes the power consumption connecting each user to the active BS with the lowest path loss if this BS still has enough capacity to support the user. Only if no other active BS can support the current end-device, then a new BS is marked as active.

The best BS locations in terms of path loss are statistically chosen after 40 simulations (step 1). The maximum number of BSs chosen will depend on the traffic demand and effective coverage per BS that guarantee at least 96% of end-devices actually covered by the network. The Path Loss (*PL)* [dB] between the end-devices and each base station (BS) is calculated as a function of the distance d [km], the frequency f [MHz], the BS antenna height hb [m], and the end-device antenna height hm [m]. For our scenario, we use the Okumura-Hata path loss model [30], which fits well with the scenario topology and related technology parameters (i.e., frequency, maximum range, and effective heights).

In the second step of the algorithm (Figure 6), the power consumption is further optimized by connecting users to the active BSs with the lowest path loss and by reducing the EIRP while the PL between the base station and end-devices is less than the maximum PL.

4. Results and Discussion

This section presents the results of the network simulations and optimizations in the considered scenario.

4.1. Sensing Network Optimization. As mentioned in Section 3.5, we calculated the optimal position for sensors using SUMO and SED tools. The main constraint of both tools is that they use approximation algorithms. As a result, we will obtain a new deployment every time we run the application (number and position of the sensors might be different) until the optimal position is found.

In order to improve the accuracy, an adaptive sampling procedure drives the selection and simulation of new samples. Modeling starts with 30 known samples using LOLA-Voronoi. Afterwards, a Kriging-based model adaptive sampling is applied. After each sampling iteration, 10 more samples are required, the received power (in dBm) for every new sample is estimated, and the process is repeated until one of the following conditions [9] is satisfied: (i) the user required accuracy has been achieved and (ii) the maximum allowed number of samples has been reached.

Once the 65 samples for each frequency (one per transmitter) are obtained, the results should be combined to obtain the final deployment, with the condition of every sensor having to receive information from the three available transmitters in the selected zone. The difference among sensors is fairly small to assume one final position for the same sensor. Only seven more sensors were required to cover the entire area. Finally, an optimal deployment of 72 sensors in the selected area was obtained. Figure 7 shows the sensors' deployment in the selected area.

Figure 8 shows the coverage zone for the three transmitters in the area. Figure 8(a) represents the coverage zone for the transmitter working on 575 MHz. It does not cover the entire zone (50 dBm of radiated power and 40 m of antenna height) because the received power in some parts of the selected area is under the TV sensitivity threshold for UHF band (-84 dBm). However, the other two transmitters transmitting on 677 MHz (Figure 8(b)) and 689 MHz (Figure 7c) located in the same building (60 dBm and 58.45 dBm, respectively, of radiated power and 145 m of antenna height) can cover the selected area without reaching the TV sensitivity threshold (-84 dBm) in any part of the area.

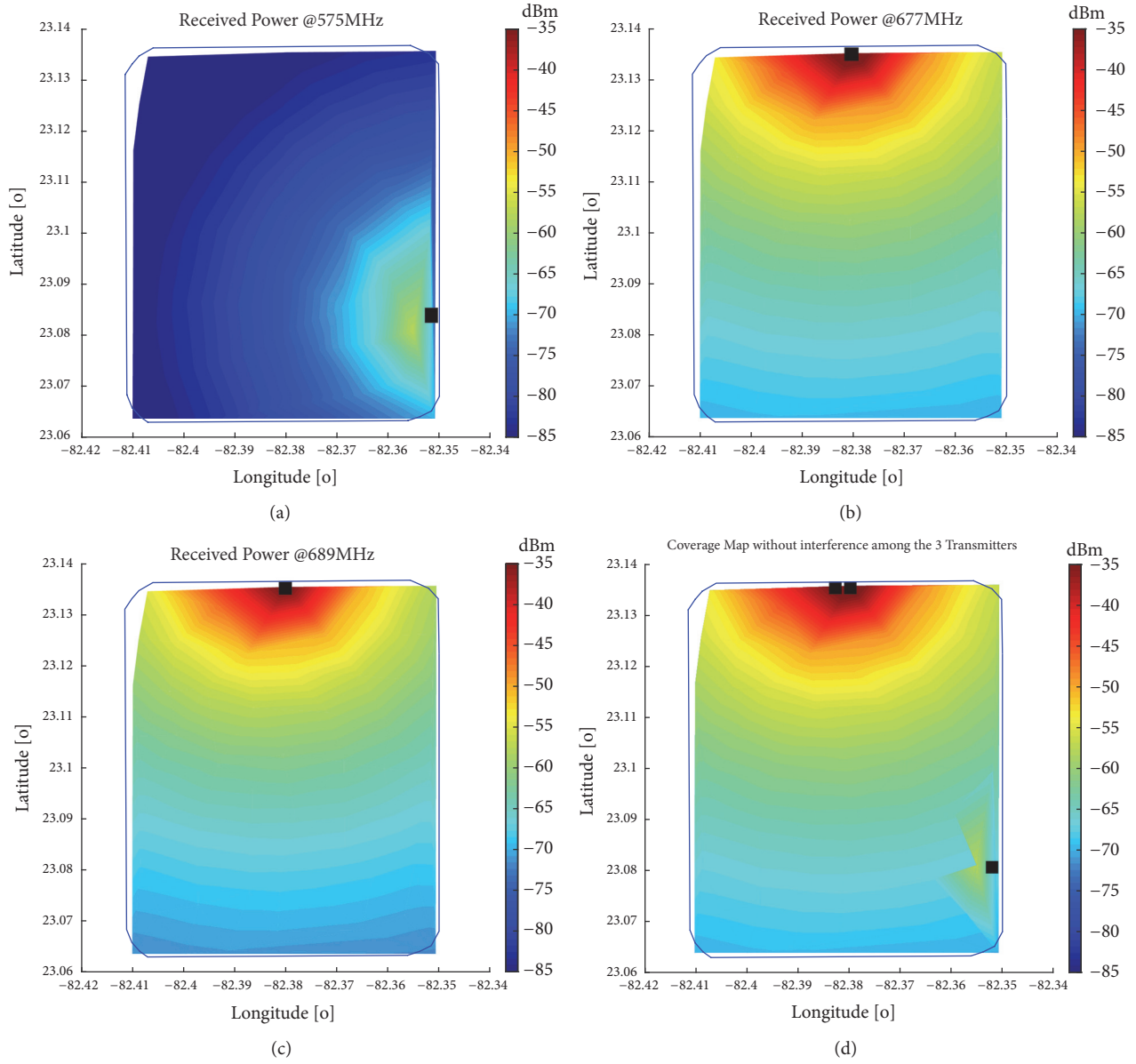


FIGURE 8: Coverage map simulation-based in the selected area, (a) 575 MHz, (b) 677 MHz, (c) 689 MHz, and (d) coverage map with the 3 transmitters (black square) working in the area without interference among them.

Every one of those transmitters is working at a different frequency and transmitting different information. Hence, there is no interference between them. Figure 8(d) shows the three transmitters together; 100 percent of the area is covered by the TV signal with a received power higher than -84 dBm in every point.

4.2. IoT Feedback Network Optimization. Figure 9 shows the two obtained solutions for the IoT feedback network coverage map in the considered area, for LoRa (Figure 9(a)) and NB-IoT (Figure 9(b)), using the method and scenarios of Section 3.2. The BSs chosen to optimally satisfy the density of connected devices and traffic are highlighted with a darker color.

The required number of BSs for LoRa is 4 and for NB-IoT is 6. Theoretically, the area coverage requirement (50 km^2) is satisfied with only 1 to 3 BSs (depending on the SF). However, a larger number of BSs are required to satisfy the capacity demand for the worst-case traffic generated by the monitoring application of the receivers. For this reason, the best LoRa network performance is achieved if all devices are capable of connecting with SF=7 to the nearest BS (in terms of path loss). The reason for this is that SF=7 achieves the highest bit rate and lowest time on air. NB-IoT has a higher capacity with QPSK modulation scheme, but the maximum coverage per BS (i.e., ~2 km) is lower than LoRa SF=7 (i.e., ~3.1 km).

Notice that the DMT will be deployed using the existing infrastructure to provide 3G and in a near feature LTE. In our

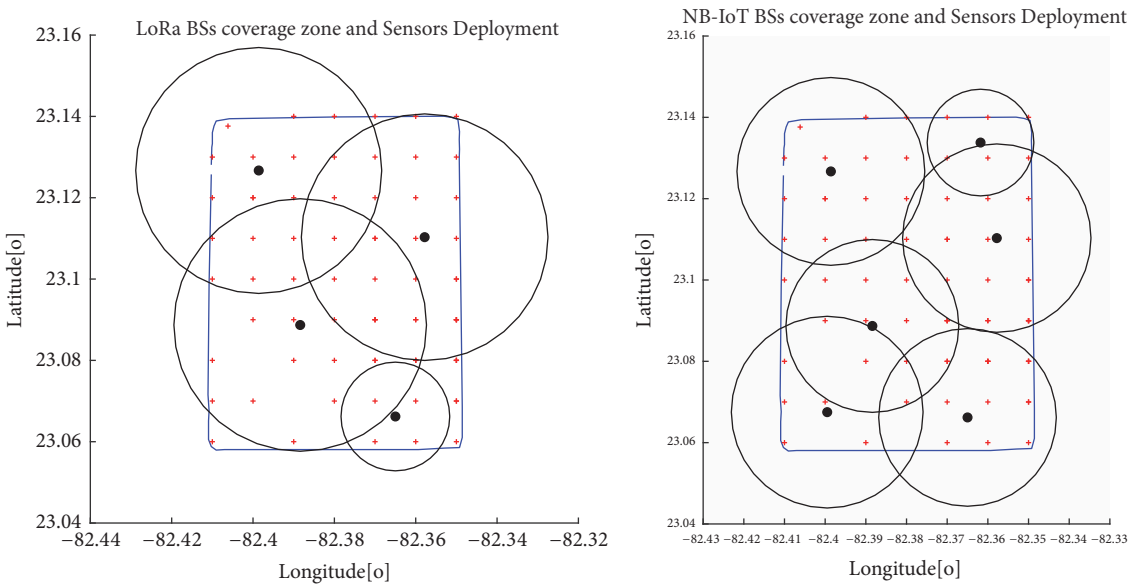


Figure 9: IoT feedback network coverage map (black circles): selected BS locations (black points) and sensors with red crosses, (a) LoRa network solution and (b) NB-IoT network solution.

solutions (LoRa and NB-IoT) more than 70% of the possible BS locations are never chosen. As such, the algorithm is a heuristic: the obtained solution might not be (nevertheless, it can be) the optimal solution, but it is a suboptimal one close to the optimal one. The approach that more than 70% of possible BSs are never chosen can be used to develop a background network to face the recovering process after a meteorological event. Removing the chosen BSs in the previous design and running the described process in Figure 6, a new optimal deployment is obtained.

4.3. Recovery Process after a Meteorological Event. The scenario presented in Section 3.2, where it is assumed that two of the three transmitters are destroyed by the hurricane, is considered. These two transmitters are located on the same building with a height of 145 m retrieving in a higher probability of the antenna destruction caused by the wind. The process to recover the original transmitters can take days and during this time, the people in the area are still uninformed about the situation they are facing.

In this case, the proposed DMT plays an important role in helping the authorities during the recovery process after a disaster event. The first step to keep the DMT running is to provide the sensing devices with batteries (Figure 1, Block I). The Raspberry PI is the core of the sensing device and it works with 5V consuming around 2A. Hence, with a power bank of 10,000 mAh, the sensing devices run around 5 hours.

The second step is to have all the necessary base stations working to collect the information from the sensing devices. In Section 4.2, we provided two solutions. The first uses 4 LoRa BSs and the second one uses 6 NB-IoT BSs. All these locations are used to provide cellular connections and are equipped with a backup power supply and cooling system, making them work during and after the hurricane. To decide the best solution, the following elements need to be accounted for: (i) access level to those locations after the hurricane, (ii) BSs' state, and (iii) in case of partial or total destruction how much time it takes to restore the service. The IoT technology will be chosen depending on the elements previously mentioned.

As mentioned, we obtained two solutions for the IoT network: 4 BSs for LoRa and 6 BSs for NB-IoT. The locations of LoRa BSs coincide in position with 4 of the 6 locations of NB-IoT BSs. However, the antennas of both technologies use the same tower. Hence, the time to repair one tower is the same no matter the technology. To have the DMT working back and save time in the recovery process, it is better to use LoRa solution.

For the scenario described in Section 3.3 where two transmitters were destroyed, the one still transmitting does not have enough power to cover the entire area (Figure 8(a)). The DMT can detect that in approximately 30% of the area the received power is under -84 dBm. In this situation and to keep the people informed, we propose using a small transmitter located in the same building with a transmission power of 40 W (around $700) instead of repairing the original one (installation time will be longer and price will be higher, around $20,000). Figure 10(a) shows the coverage area of the new transmitter (~46 dBm, 145 m of antenna height). Figure 10(b) shows the two transmitters together and 90% of the area is covered by the TV signal with a signal strength higher than -84 dBm. Only 10% (black square) of the area still remains as an unserved area.

In Section 3.3, we assumed that 60% of the 41 available base stations we needed to provide a cellular connection

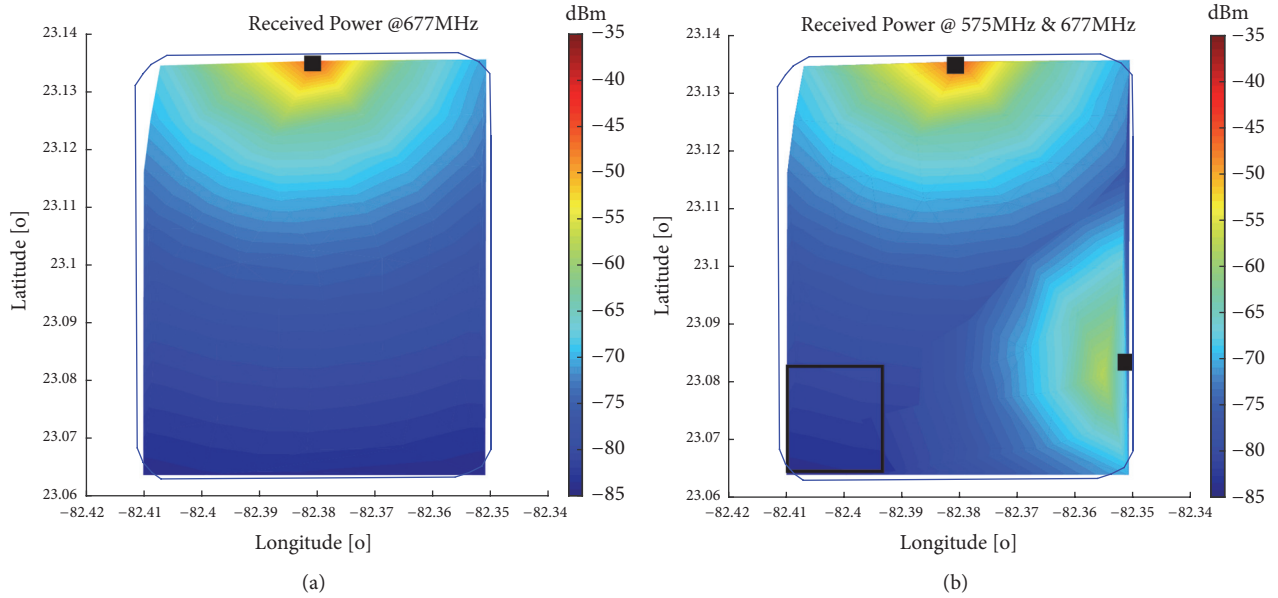


FIGURE 10: Coverage map (received-power) simulation-based in the selected area, (a) 677 MHz with a transmitted power of 40 W, (b) coverage map with the 2 transmitters (black square) working in the area without interference between them. The black square is the unserved area, around 10% of the total surface.

(GSM, 3G, and 4G) were destroyed by the hurricane (25 BSs destroyed). After a hurricane, it is difficult to check all the BSs in a short time period. In this situation, the proposed DMT is useful to precisely detect which BS is working and give the authorities feedback about the real situation.

4.4. Network Deployment Cost Considerations. For network deployment, we need to consider the end-devices structure and the BS infrastructure. A single sensing device is built with an SDRplay sensor (model RSP1A), a Raspberry PI 3 (model B+), and an IoT transceiver (LoRa and NB-IoT). The approximate price for all of these devices is $215. The total price for 72 sensors is approximately $15,480. A single LoRa BS has deployment cost up to $1,000 and a single NB-IoT of $15,000, considering the reusage of LTE infrastructure. The difference in infrastructure is not significant (2 BSs more for NB-IoT). However, the NB-IoT license cost for the mobile operator is thus considerably higher. The total LoRa BSs infrastructure cost is just $4,000 (4 BSs). For NB-IoT, the total BSs deployment cost is $90,000 (approximately 22 times higher for 6 BSs).

5. Conclusion

In this paper, we investigated the feasibility of building a REM deploying 72 SDR sensors using SUMO and SED tool in a selected area and using LoRa and/or NB-IoT technologies for networking. The IoT feedback network is designed and optimized for minimal power consumption and infrastructure.

This DMT permits the regulatory body to check the correct use of the assigned spectrum and plays an important role in the recovery process after a catastrophic event, such as a hurricane, where radio and TV broadcasting are important in keeping people informed. Besides, the tool can be used to deploy new services in the nonserved area (e.g., 4G in the 700 MHz).

Future research will consist of the emulation of the DMT in a real scenario.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

M. Deruyck is a Postdoctoral Fellow of the FWO-V (Research Foundation, Flanders, Belgium).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Y. Hervis Santana is supported by *LACETEL* and a doctoral grant from the Special Research Fund (BOF) of Ghent University, Belgium.

References

[1] R. M. Alonso et al., "IoT-Based Management Platform for Real-Time Spectrum and Energy Optimization of Broadcasting

Networks," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

[2] V. Popescu, M. Fadda, M. Murroni, J. Morgade, and P. Angueira, "Co-channel and adjacent channel interference and protection issues for DVB-T2 and IEEE 802.22 WRAN operation," *IEEE Transactions on Broadcasting*, vol. 60, no. 4, pp. 693–700, 2014.

[3] M. Kosunen, V. Turunen, K. Kokkinen, and J. Ryynanen, "Survey and Analysis of Cyclostationary Signal Detector Implementations on FPGA," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 3, no. 4, pp. 541–551, 2013.

[4] Z. Quan, S. Cui, and A. H. Sayed, "Optimal Linear Cooperation for Spectrum Sensing in Cognitive Radio Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 28–40, 2008.

[5] D. Sinha, A. K. Verma, and S. Kumar, "Software defined radio: Operation, challenges and possible solutions," in *Proceedings of the 2016 10th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1–5, 2016.

[6] M. B. Sruthi, M. Abirami, A. Manikkoth, R. Gandhiraj, and K. P. Soman, "Low cost digital transceiver design for software defined radio using RTL-SDR," in *Proceedings of the 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, pp. 852–855, 2013.

[7] N. LLC, "Nuand — bladeRF Software Defined Radio," http://www.nuand.com, Accessed: 03-Dec-2018.

[8] L. De Tommasi, D. Gorissen, J. A. Croon, and T. Dhaene, "Surrogate Modeling of RF Circuit Blocks," *Mathematics in Industry*, vol. 15, pp. 2–7, 2010.

[9] L. De Tommasi, D. Gorissen, J. Croon, and T. Dhaene, "Surrogate Modeling of Low Noise Amplifiers Based on Transistor Level Simulations," *Scientific Computing in Electrical Engineering*, vol. 14, pp. 225–232, 2010.

[10] K. Crombecq, I. Couckuyt, D. Gorissen, and T. Dhaene, "Space-filling sequential design strategies for adaptive surrogate modelling," in *Proceedings of the First International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering, B.H.V*, 2009.

[11] K. Crombecq, L. De Tommasi, D. Gorissen, and T. Dhaene, "A novel sequential design strategy for global surrogate modeling," in *Proceedings of the 2009 Winter Simulation Conference, WSC*, pp. 731–742, 2009.

[12] S. Srinu, S. L. Sabat, and S. K. Udgata, "FPGA implementation of cooperative spectrum sensing for cognitive radio networks," in *Proceedings of the 2010 Second UK-India-IDRC International Workshop on Cognitive Wireless Systems (UKIWCWS)*, pp. 1–5, 2010.

[13] A. Sahai, N. Hoven, and R. Tandra, "Some Fundamental Limits on Cognitive Radio," in *Proceedings of the Allert. Conf. Control. Commun. Comput*, pp. 1662–1671, 2004.

[14] H. Birkan Yilmaz, T. Tugcu, and F. Alagoz, "Novel quantization-based spectrum sensing scheme under imperfect reporting channel and false reports," *International Journal of Communication Systems*, vol. 27, no. 10, pp. 1459–1475, 2014.

[15] K. Crombecq, *Surrogate Modeling of Computer Experiments with Sequential Experimental Design*, Universiteit Antwerpen, 2011.

[16] R. Martin and R. Thomas, "Algorithms and bounds for estimating location, directionality, and environmental parameters of primary spectrum users," *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5692–5701, 2009.

[17] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1847–1862, 2010.

[18] Y. Zhao, J. Gaeddert, K. K. Bae, and J. H. Reed, "Radio Environment Map Enabled Situation-Aware Cognitive Radio Learning Algorithms," in *Proceedings of the Tech. Conf. Prod. Expo*, 2006.

[19] H. Yilmaz Birkan, T. Tugcu, F. Alagöz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162–169, 2013.

[20] R. S. Sinha, Y. Wei, and S.-H. Hwang, "A survey on LPWA technology: LoRa and NB-IoT," *ScienceDirect*, vol. 3, no. 1, pp. 14–21, 2017.

[21] Y. E. Wang, X. Lin, A. Adhikary et al., "A Primer on 3GPP Narrowband Internet of Things," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117–123, 2017.

[22] D. Ball, N. Naik, and P. Jenkins, "Lightweight and Cost-Effective Spectrum Analyser Based on Software Defined Radio and Raspberry Pi," in *Proceedings of the 2017 European Modelling Symposium (EMS)*, pp. 260–266, 2017.

[23] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the Limits of LoRaWAN," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017.

[24] M. Deruyck, E. Tanghe, W. Joseph, and L. Martens, "Modelling and optimization of power consumption in wireless access networks," *Computer Communications*, vol. 34, no. 17, pp. 2036–2046, 2011.

[25] "S. A. of the P. Republic and China, "GB20600-2006 Standard: Framing Structure, Channel Coding and Modulation for Digital Television Terrestrial Broadcasting System (DTMB),"," 2006.

[26] "Cuba después del huracán Irma: Últimas noticias y testimonios III (miércoles 13 de septiembre) — Cubadebate," http://www.cubadebate.cu/temas/politica-temas/2017/09/13/cuba-despues-del-huracan-irma-ultimas-noticias-y-testimonios-iii/#.XAZ-_GhKi-Uk, 2017, Accessed: 04-Dec-2018.

[27] T. Rheinland, *A technical overview of LoRa® and LoRaWAN™ What is it?*, 2015.

[28] 3GPP, *Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things*, Valbonne-France, 2015, Release 13.

[29] L. Zhang, A. Ijaz, P. Xiao, and R. Tafazolli, "Channel Equalization and Interference Analysis for Uplink Narrowband Internet of Things (NB-IoT)," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2206–2209, 2017.

[30] A. F. Molish and Fellow IEEE, *Wireless Communications*, Calif, USA, 2nd edition, 2011.

[31] "Mi. de Comunicaciones," RESOLUCIÓN No. 124 /2016. Cuba, 2016, p. 11.

[32] D. Gorissen, I. Couckuyt, P. Dhaene, and T. Demeester, "A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based," 2010.

[33] J. van der Herten, I. Couckuyt, D. Deschrijver, and T. Dhaene, "Adaptive classification under computational budget constraints using sequential data gathering," *Advances in Engineering Software*, vol. 99, pp. 137–146, 2016.

[34] D. Deschrijver, T. Dhaene, and J. Broeckhove, "Adaptive model based parameter estimation, based on sparse data and frequency derivatives," in *Proceedings of the International Conference on Computational Science*, pp. 443–450, 2004.

[35] W. C. M. Van Beers, "Kriging metamodeling in discrete-event simulation: An overview," in *Proceedings of the 2005 Winter Simulation Conference*, p. 2, 2005.

[36] L. Balewski and M. Mrozowski, "Creating neural models using an adaptive algorithm for optimal size of neural network and training set," in *Proceedings of the 15th International Conference on Microwaves, Radar and Wireless Communications*, vol. 2, pp. 543–546, IEEE Cat. No.04EX824.

[37] A. Lamecki, P. Kozakowski, and M. Mrozowski, "CAD-model construction based on adaptive radial basis functions interpolation technique," in *Proceedings of the 15th International Conference on Microwaves, Radar and Wireless Communications, MIKON - 2004*, pp. 799–802, IEEE Cat. No.04EX824, 2004.

[38] H. Zhao, D. Knight, E. Taskinoglu, and V. Jovanovic, "Data driven design optimization methodology development and application," in *Proceedings of the International Conference on Computational Science (LNCS 3038)*, pp. 748–755, 2004.

[39] K. Crombecq and T. Dhaene, *Generating Sequential Space-filling Designs Using Genetic Algorithms and Monte Carlo Methods*, 2010.

[40] K. Crombecq, E. Laermans, and T. Dhaene, "Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling," *European Journal of Operational Research*, vol. 214, no. 3, pp. 683–696, 2011.

[41] M. Deruyck, J. Wyckmans, L. Martens, and W. Joseph, "Emergency ad-hoc networks by using drone mounted base stations for a disaster scenario," in *Proceedings of the 12th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob*, pp. 1–7, 2016.

[42] R. Martinez Alonso, D. Plets, M. Deruyck, L. Martens, G. Guillen Nieto, and W. Joseph, "TV white space and LTE network optimization toward energy efficiency in suburban and rural scenarios," *IEEE Transactions on Broadcasting*, vol. 64, no. 1, pp. 164–171, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7080249, 11 pages
https://doi.org/10.1155/2019/7080249

Research Article

Sensing Coverage Algorithm of Sparse Mobile Sensor Node with Trade-Off between Packet Loss Rate and Transmission Delay

Kehua Zhao,[1] **Yourong Chen**,[1,2] **Siyi Lu**,[2] **Banteng Liu**,[1] **Tiaojuan Ren**,[1] **and Zhangquan Wang**[1]

[1]*College of Information Science and Technology, Zhejiang Shuren University, Hangzhou, Zhejiang 310015, China*
[2]*School of Information Science & Engineering, Changzhou University, Changzhou, Jiangshu 213164, China*

Correspondence should be addressed to Yourong Chen; jack_chenyr@163.com

Received 25 March 2019; Revised 16 June 2019; Accepted 23 June 2019; Published 8 July 2019

Guest Editor: Maurizio Casoni

To solve the problem of sensing coverage of sparse wireless sensor networks, the movement of sensor nodes is considered and a sensing coverage algorithm of sparse mobile sensor node with trade-off between packet loss rate and transmission delay (SCA_SM) is proposed. Firstly, SCA_SM divides the monitoring area into several grids of same size and establishes a path planning model of multisensor nodes' movement. Secondly, the social foraging behavior of Escherichia coli in bacterial foraging is used. A fitness function formula of sensor nodes' moving paths is proposed. The optimal moving paths of all mobile sensor nodes which can cover the entire monitoring area are obtained through the operations of chemotaxis, replication, and migration. The simulation results show that SCA_SM can fully cover the monitoring area and reduce the packet loss rate and data transmission delay in the process of data transmission. Under certain conditions, SCA_SM is better than RAND_D, HILBERT, and TCM.

1. Introduction

Wireless sensor networks (WSNs) are composed of a large number of sensor nodes and are deployed manually or randomly in a given monitoring area. The sensor node recognizes and collects the state of the monitoring environment (such as temperature, humidity, and brightness) and sends this information to devices (sink nodes or base stations) that are suitable for data processing, visualization, analysis, and decision-making [1, 2]. Because sensing range and energy of sensor nodes are limited, the analysis of large amounts of local data is impossible when the monitoring area is huge and the environment is complex. Moreover, large-scale deployment requires a huge cost investment. Therefore, a sparse mobile sensor network with a small number of sensor nodes is considered; in particular, a device such as a drone or robot can be used in the sensor node to move to each monitoring position. The mobile monitoring node can sense the entire monitoring area. It reduces the system application cost [3].

However, sparse mobile sensor network may generate a large packet loss rate and a long data transmission delay. Therefore, mobile path selection and sensing coverage of mobile sensor nodes need be studied. At present, some scholars focus on the path selection of mobile sensor nodes installed on UAVs (unmanned aerial vehicle) and achieve certain results. For example, a rasterized motion model is established based on the environmental information within the scope of the UAV in the literature [4], and A$*$ algorithm is used to obtain the UAV's moving path. The literature [5] constructs the model of UAV flight path, quantitatively describes the threat area on the path, and uses improved artificial potential field method to obtain a sufficiently smooth flight path. A three-layer framework consisting of sensor layer, cluster head layer, and moving collector layer is proposed in the literature [6]. The mobile collector calculates its moving

trajectory planning, uses the divide and conquer method to determine the turning point on the path, and obtains its efficient mobile data collection scheme. In the literature [7], each mobile sensor node is assigned a unique ID to find its potential moving position. The greedy algorithm is used to select the optimal stayed position from these potential moving positions. Finally, its optimal path is constructed. The literatures [4–7] mainly implement the path of mobile sensor node from starting position to target position, but how to efficiently cover the entire monitoring area is not taken into consideration. The literature [8] uses HILBERT curve as moving path of mobile aggregation node to cover the entire network. The literature [9] proposes a temporal coverage mechanism (TCM) for mobile sensor nodes in a sparse network environment. TCM divides the monitoring area into several grids so that mobile sensor nodes can cover the entire monitoring area as quickly as possible. Meanwhile, to achieve efficient coverage of the monitoring area under the sparse WSNs, some scholars focus on using artificial intelligence algorithms to obtain the moving paths of mobile sensor nodes. For example, the literature [10] proposes an improved artificial bee colony algorithm to solve the traveling salesman problem of neighborhood search in sparse WSNs and obtains the optimal data collection and movement paths of mobile sensor nodes. The literature [11] proposes autonomous movement model of mobile sensor node, which can be solved by bacterial foraging optimization algorithm (BFOA). The autonomous mobile solution of each mobile sensor node can be obtained. However, the literatures [8–11] do not consider that the storage space of mobile sensor nodes is limited and have large packet loss rate and high data transmission delay.

Therefore, a sensing coverage algorithm of sparse mobile sensor node with trade-off between packet loss rate and transmission delay (SCA_SM) is proposed. The monitoring area is divided into a finite number of square virtual grids by SCA_SM. Each mobile sensor node takes neighbor grid center as next stayed position, proposes a mathematical expression of moving paths of mobile sensor nodes, and establishes a moving path planning model for multiple sensor nodes. To quickly solve the optimization model, the processes of chemotaxis, replication, and migration in the BFOA are used to solve the model. In accordance with the optimization goal of the WSNs, fitness function is proposed to reduce the packet loss rate and data transmission delay as much as possible. Mobile sensor node simulates two basic actions of Escherichia coli, such as advancement and flipping in the search space. The mobile sensor node first moves a step in any random direction. If the fitness value of current moving path in this direction is improved compared with that value of previous moving path, then it will continue to advance in this direction. Instead, it will flip to find a new random direction and a position that can improve the fitness value. When all mobile sensor nodes cover the entire monitoring area or reach the end condition, SCA_SM outputs an optimal solution. Therefore, optimal moving paths of all mobile sensor nodes which cover the entire monitoring area are obtained, and the packet loss rate and data transmission delay in data transmission process are reduced.

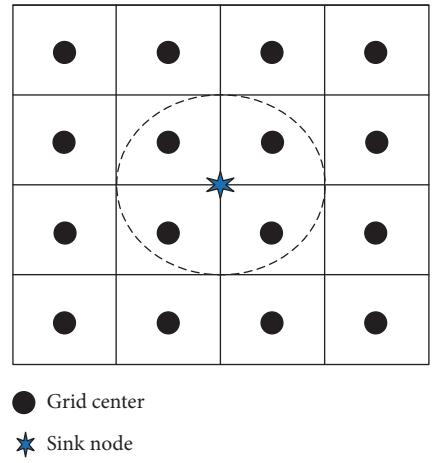


Figure 1: Principle of SCA_SM.

2. Algorithm Assumptions and Basic Principles

In SCA_SM, WSNs are assumed to be evenly distributed in the monitoring area of two dimensions. Multiple mobile sensor nodes and one static sink node exist in the network. Mobile sensor node has a fixed communication radius and data storage capacity. The position coordinates of the nodes are obtained by satellite positioning modules, such as GPS, Beidou, or other positioning modules. Only one grid center is allowed to pass once in one round of data collection. Static sink node collects data at the center of the monitoring area and can only listen to the data reported by mobile sensor nodes in the single communication range of sink node.

As shown in Figure 1, the monitoring area is divided into square grids of uniform size, and some mobile sensor nodes are randomly placed in the monitoring area. When the network is running, all mobile sensor nodes move to the neighbor grid center's positions which are not stayed before and sense the data of grid area. If a mobile sensor node is within the communication range of sink node, then the data are sent to sink node in single-hop manner; otherwise, the data are stored in the mobile sensor node. If storage space of mobile sensor node is full, the oldest data are deleted, and the received data are stored. However, SCA_SM still needs to solve the following two problems. The first is how to establish an optimization model of multisensor nodes' mobile sensing coverage through mathematical model. The second is how to use the improved bacterial foraging algorithm to solve the optimization model and obtain optimal moving path scheme of multiple mobile sensor nodes that can fully cover the monitoring area. The specific solutions to these two problems are as follows.

3. Establishment of the Optimization Model

The optimization model of multisensor nodes' mobile sensing coverage can be converted into

$$\min \quad \left(x_1 D_R + \frac{x_2 T_{average}}{T_{yu}} \right) \tag{1}$$

$$\text{s.t.} \quad P^k = \left\{ p_1^k, p_2^k, p_3^k, \ldots, p_j^k \right\} \tag{1a}$$

$$p_j^k \neq p_i^k, \quad \forall j \text{ and } j \neq i, \ \forall k \tag{1b}$$

$$C_{t+1}^k = \begin{cases} C_{\max} & \left(C_t^k + SD_{t+1}^k \right) \geq C_{\max} \text{ and } d_{ks} > d_{\max} \\ C_t^k + SD_{t+1}^k & \left(C_t^k + SD_{t+1}^k \right) < C_{\max} \text{ and } d_{ks} > d_{\max} \\ 0 & d_{ks} \leq d_{\max} \end{cases}, \quad \forall k \tag{1c}$$

$$D_{t+1}^k = \begin{cases} D_t^k + C_t^k + SD_{t+1}^k - C_{\max}, & \left(C_t^k + SD_{t+1} \right) \geq C_{\max} \text{ and } d_{ks} > d_{\max} \\ D_t^k & d_{ks} \leq d_{\max} \end{cases}, \quad \forall k \tag{1d}$$

$$\cup Cover_k = 1, \quad \forall k \tag{1e}$$

$$D_R = \frac{\sum_k D_{t+1}^k}{N_{total}} \tag{1f}$$

$$T_{average} = \frac{\left(\sum_m \left(T_m - t_m \right) + \sum_n \left(t - t_n \right) \right)}{N_{pack}} \tag{1g}$$

$$x_1 + x_2 = 1 \tag{1h}$$

where P^k represents the moving path of mobile sensor node k. p_j^k represents the jth stayed grid center's position of mobile sensor node k. C_t^k represents the amount of data stored in the mobile sensor node k at time t. $C_{\max}$ represents the maximum storage space of mobile sensor node. d_{ks} represents the distance from mobile sensor node k to sink node. $d_{\max}$ represents the maximum communication radius of sink node. SD_{t+1}^k represents the amount of data received and sensed by mobile sensor node k at the next moment. D_t^k represents the total number of lost packets of mobile sensor node k from network start time to time t. $Cover_k$ represents the sensing coverage of mobile sensor node k, that is, the ratio of the number of grid center positions in the moving path of mobile sensor node k and the total number of grids. D_R represents the packet loss rate. N_{total} represents the total number of packets generated by all mobile sensor nodes. $T_{average}$ represents the current data transmission delay. T_m represents the time when the packet m is successfully sent to sink node. t_m represents the generation time of packet m. t_n represents the generation time of packet n stored in mobile sensor node or discarded. N_{pack} represents the total number of packets generated by mobile sensor nodes after time t. T_{yu} represents the threshold of data transmission delay. x_1 represents the weight factor of packet loss rate. x_2 represents the weight factor of data transmission delay.

Equation (1) represents the objective function of optimization model, namely, minimizing the packet loss rate and data transmission delay. Formula (1a) represents that the moving path of mobile sensor node k is composed positions of grid centers. Formula (1b) represents that each mobile sensor node does not repeat through the same grid center during one round of movement. Formula (1c) represents that when mobile sensor node k is within the communication range of sink node, the data are directly sent to sink node; otherwise, the data are stored in the cache space. If the cache space is full, then the oldest data will be discarded. Formula (1d) represents an updated calculation formula for the total number of lost packets of mobile sensor node. Formula (1e) represents that the moving paths of all mobile sensor nodes contain all grid centers; namely, the sensing coverage equals 1. Formula (1f) is a formula for calculating the packet loss rate. Formula (1g) is a formula for calculating the data transmission delay.

4. Algorithm Solving

BFOA is a distributed parallel and random global optimization algorithm that is widely used in image processing, machine learning, pattern recognition, workshop scheduling, and other fields [12]. BFOA simulates the foraging behavior of Escherichia coli in human intestines and seeks optimal population through four operations: chemotaxis, aggregation, replication, and migration. The solving process for the optimal moving paths of mobile sensor nodes is the process of Escherichia coli searching for the area of abundant food. However, the solution of bacterial fitness value needs to be solved and BFOA needs to be improved

in the solution process. Therefore, each bacterium is moving paths of all mobile sensor nodes which sense and cover the entire monitoring area. $data(S_x, S_y, k, t)$ represents the data sensed by mobile sensor node k on the grid (S_x, S_y) at time t. L_{now} represents the length of current moving path of mobile sensor node. G_{rest} represents the number of grids that mobile sensor node has not passed. $Cover_{sum}$ represents the sum of coverage of all mobile sensor nodes. N_m represents the total number of mobile sensor nodes. S represents the size of bacterial population. N_{ed}, N_{re}, and N_c represent the number of migration, replication, and chemotaxis, respectively. P_{ed} represents the probability of migration. N_m represents the number of mobile sensor nodes. $J(i, g, h, l)$ represents fitness value of bacterium i at the gth chemotaxis, hth replicas and lth migration. NG_i^k represents the neighbor grid set of mobile sensor node k in the bacterium i. $S^i = \{S_1^i, S_2^i, S_3^i, ..., S_k^i\}$ represents the movement scheme of current bacterium i after migration, and it is initially an empty set. S_k^i represents the moving path of mobile sensor node k in the bacterium i. The specific content of the model solution is as follows.

4.1. Calculation of Bacterial Fitness Value. Because the setting of bacterial fitness value affects the solution of optimization model (1), the chemotaxis operation leads the bacteria to preferentially move to a nutrient-rich environment. But if the fitness value of the bacterium i is directly calculated by model (1), mobile sensor node can easily preferentially move to the grid within the communication range of sink node. It results in a large packet loss rate and data transmission delay. Therefore, different fitness value formulas are adopted in the actual path finding process and after the path is searched. The specific implementation steps of bacterial fitness value calculation are as follows.

Step 1. Initialize the parameters and the storage space of all mobile sensor nodes. Initialize the storage space of sink node. t=1, k=1.

Step 2. At the current time t, the data of mobile sensor node k at the position (S_x, S_y) are sensed, stored, and recorded as $data(S_x, S_y, t)$. The data storage amount C_{t+1}^k at next moment according to formula (1c) is updated. If the data storage amount C_{t+1}^k at next moment is greater than maximum storage space $SC_{\max}$, then the oldest data are deleted. According to formula (1d), the deleted data are recorded in the number of lost packets, and the number of lost packets is updated.

Step 3 (t=t+1). Mobile sensor node k moves to next stayed position. If mobile sensor node k is within the single-hop range of sink node, all data in the storage space are directly sent to sink node. The data received by sink node are recorded, and the data storage capacity is updated. Then whether a neighbor mobile sensor node exists is determined. If it exists, then it communicates with the neighbor mobile sensor node, obtains coverage rate of the neighbor sensor node, and calculates the sum $Cover_{sum}$ of sensing coverage of the known mobile sensor nodes.

Step 4. If $t \le L_{now}$, then skip to Step 2; otherwise, k=k+1, t=1. If $k \le N_m$, then skip to Step 2; otherwise, skip to Step 5.

Step 5. If $Cover_{sum}$=1, then packet loss rate D_R is calculated by formula (1f). Skip to Step 7; otherwise, the mobile sensor node is still looking for the path. The number S_{num} of grids in the single-hop communication range of sink node is calculated. The number N_{loss} of discarded packets is calculated by formula (2).

$$N_{loss} = G_{rest} + C_t - S_{num}\left(C_{\max} + 1\right) \tag{2}$$

where G_{rest} represents the number of grids which is not passed before. C_t represents the sum of data stored by all mobile sensor nodes.

Step 6. When $N_{loss} \le 0$, the grids in the single-hop communication range of sink node is allocated reasonably, and the other grids do not generate the number of lost packets, $N_{loss} = 0$. Otherwise, N_{loss} is recorded in the number of lost packets, and the packet loss rate D_R' is calculated by formula (3) on the basis of received data packets of sink node and the number of lost packets stored in mobile sensor nodes.

$$D_R' = \frac{\left(N_t + N_{loss}\right)}{N_{total}} \tag{3}$$

where N_t represents number of lost packets of mobile sensor node at previous moment and N_{total} represents total number of packets generated by all mobile sensor nodes.

Step 7. The estimate value $T_{average}$ of data transmission delay is calculated by formula (1g), and the fitness values of all bacteria are calculated by formula (4).

$$fitness = x_1 \overline{D}_R + \frac{x_2 T_{average}}{T_{yu}} \tag{4}$$

where packet loss rate $\overline{D}_R$ is

$$\overline{D}_R = \begin{cases} D_R', & Cover_{sum} < 1 \\ D_R, & Cover_{sum} = 1. \end{cases} \tag{5}$$

4.2. Chemotaxis of Bacteria. Mobile sensor node simulates two basic actions of Escherichia coli, such as advancement and flipping in the path planning process. First, each mobile sensor node in a bacterium moves a step in any random direction. If the fitness value of current moving path in this direction is improved compared with the fitness value of previous moving path, then it will continue to advance in this direction. Otherwise, it will flip to find a new random direction and a position that can improve the fitness value. If neighbor grid has been accessed, the mobile sensor node is easy to fall into dead end. At this time, the nearest grid should be selected to meet the constraint condition (1b), thereby achieving the chemotaxis operation. As shown in Figure 2, the specific implementation steps are as follows.

Step 1. Obtain all current bacterial information and replication parameter h, and let g=1, l=1, i=1, k=1.

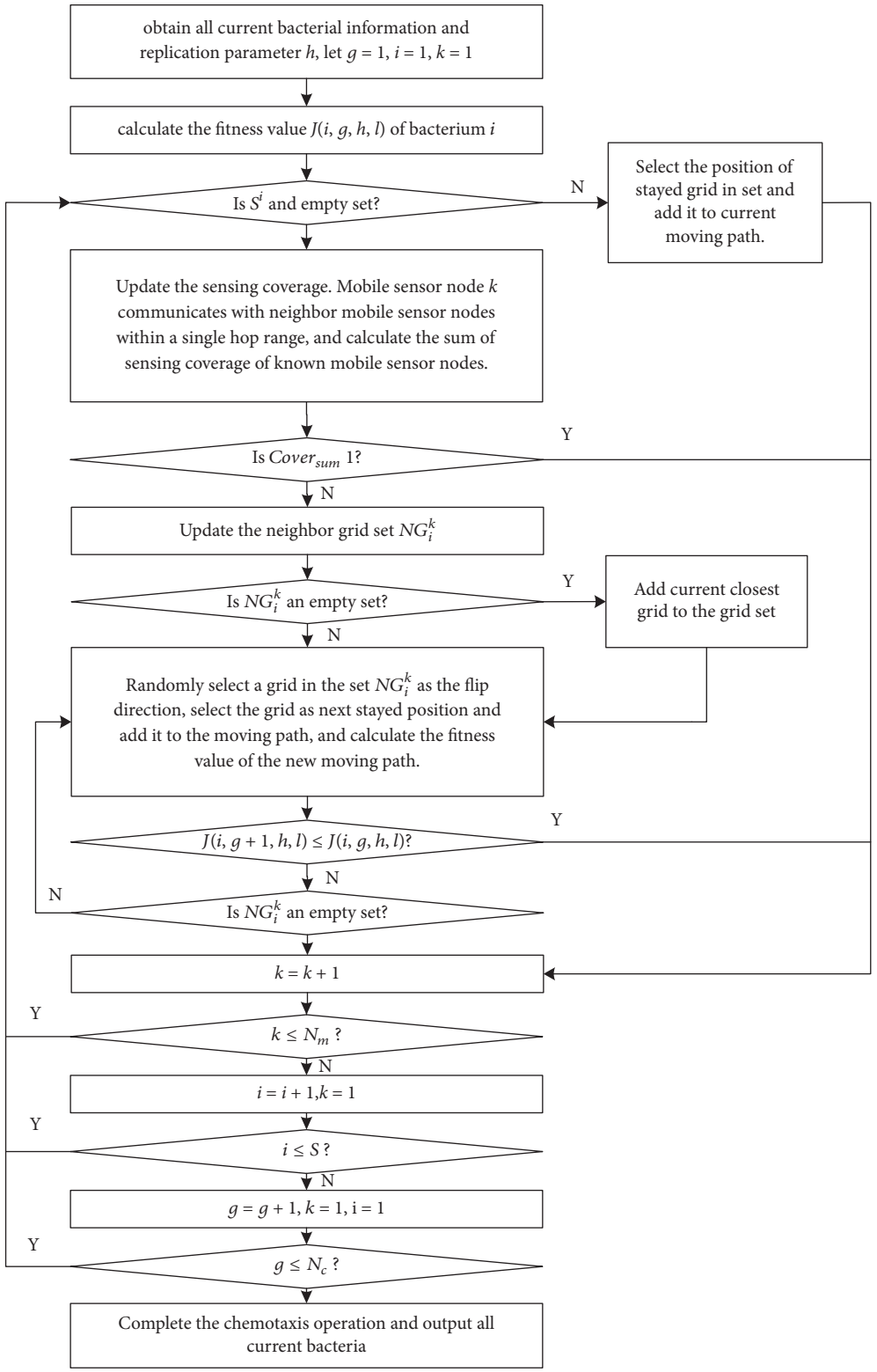


Figure 2: Flow chart of chemotaxis.

Step 2. Calculate the fitness value $J(i, g, h, l)$ of the bacterium i. It is judged whether the set S^i is an empty set. If it is an empty set, then skip to Step 3. Otherwise, let $S_g^i = g + 1 + N_c(h - 1)$. Select the S_g^ith stayed grid in the set S_k^i. Let it be the next grid of mobile sensor node k. Add it to current moving path, and obtain the new moving path. Skip to Step 6.

Step 3. Update the sensing coverage $Cover_k$, and mobile sensor node k communicates with neighbor mobile sensor nodes in the single-hop range to calculate the sum $Cover_{sum}$ of sensing coverage of known mobile sensor nodes. If $Cover_{sum}$ is 1, then skip to Step 6; otherwise, skip to Step 4.

Step 4. Update the neighbor grid set NG_i^k, and determine whether the set NG_i^k is an empty set. If it is, add current closest gird to the grid set NG_i^k; otherwise, neighbor grid set NG_i^k is not updated. Randomly select a grid in the set NG_i^k as the flip direction, delete the grid in the set NG_i^k, add the next stayed grid to moving path, and calculate the fitness value $J(i, g + 1, h, l)$ of the new moving path.

Step 5. If $J(i, g + 1, h, l) \leq J(i, g, h, l)$, it indicates that the bacteria are closer to optimal solution, so skip to Step 6. Otherwise, if fitness value is not improved, then judge whether the set NG_i^k is an empty set. If it is, skip to Step 6; otherwise, skip to Step 4 and reselect the random direction for chemotaxis.

Step 6 (k=k+1). If $k \leq N_m$, then skip to Step 2; otherwise, $i = i + 1$, k=1. If $i \leq S$, then skip to Step 2; otherwise, g=g+1, k=1, i=1. If $g \leq N_c$, then skip to Step 2; otherwise, the chemotaxis operation is completed.

4.3. Improved BFOA. As shown in Figure 3, SCA_SM calculates bacterial fitness value through formula (4) and uses chemotaxis, replication, and migration of improved bacterial foraging algorithm to solve the model (1). The specific implementation steps are as follows.

Step 1. Initialize the bacterial population and algorithm parameters: the initial positions of all mobile sensor nodes in bacteria are randomly generated, and neighbor grid set is updated. Let F_{best} be infinite, g=1, h=1, l=1, i=1, k=1.

Step 2. Use chemotaxis to obtain moving paths of mobile sensor nodes in all bacteria.

Step 3. Calculate the fitness values of S bacteria in the population and arrange them in ascending order. Remove half of the bacteria with high fitness value and copy half of the bacteria with small fitness value. h=h+1. If $h \leq N_{re}$, i=1, g=1, then skip to Step 2; otherwise, skip to next step.

Step 4. Select the moving paths and fitness value of the bacterium with smallest fitness value F_{local}. If $F_{local} < F_{best}$, $F_{best} = F_{local}$, then the moving paths of mobile sensor nodes in the bacterium are the optimal moving paths. Otherwise, the bacterium is eliminated and the optimal moving paths of mobile sensor nodes are replicated to the bacterium.

Step 5. Calculate the adaptive probabilities P_{self}^i of all S bacteria via formula (6).

$$P_{self}^i = \frac{J_{num}^i - J_{num}^{\min}}{J_{num}^{\max} - J_{num}^{\min}} P_{ed} \tag{6}$$

where J_{num} represents the fitness function and P_{ed} represents the migration probability.

Step 6. Generate a random number between 0 and 1. If the adaptive probability P_{self}^i is greater than the random number, initial positions of N_m mobile sensor nodes are randomly generated as current grid positions and current moving paths in the bacterium i. Let the set S^i be an empty set; otherwise, the moving paths of bacterium i are saved in set S^i.

Step 7 (l=l+1). If $l \leq N_{ed}$, then skip to Step 2; otherwise, output the optimal moving paths of all mobile sensor nodes and end the algorithm.

5. Algorithm Simulation

5.1. Simulation Parameter Selection. To verify the algorithm performance, the simulation experiment is performed and SCA_SM is compared with RAND_D, HILBERT [8], and TCM [9]. Assume that the 200 m∗200 m monitoring area is divided into several virtual square grids. The sink node is in the center of the monitoring area. The positions of mobile sensor nodes are random, and the communication radius of each nodes is 50 m. In the simulation, the main parameters are set to the following values. The maximum storage space $C_{\max}$ is 6 kbit. The amount of data received and sensed by the node in each uncovered grid is 2 kbit. The threshold T_{yu} of data transmission delay is 10^2. The weight factor x_1 of packet loss rate and weight factor x_2 of data transmission delay are 0.5. The chemotaxis number N_c is 10; the replication number N_{re} is 20. The migration number N_{ed} is 50. The migration probability P_{ed} is 0.6. The number of bacteria S is 50 when square grid number is less than or equal to 150, and the number of bacteria S is 100 when square grid number is greater than 150.

5.2. Analysis of Simulation Results. Because multiple mobile sensor nodes are considered and randomly placed within the monitoring area, initial positions of mobile sensor nodes are randomly generated. To illustrate the effectiveness of the algorithms, random initial positions with (25, 25), (25, 175), and (175, 175) are selected, and the number of mobile sensor nodes is 2, 3, and 4. Then the parameters in Section 4.1 are selected and Figures 4–7 are obtained. As shown in Figure 4, all mobile sensor nodes in RAND_D randomly select the closest grid which is not stayed before as next stayed position. As shown in Figure 5, the number of stayed positions of HILBERT is a power of 4, and each mobile sensor node moves following HILBERT path until the entire monitoring area is covered. As shown in Figure 6, each mobile sensor node in

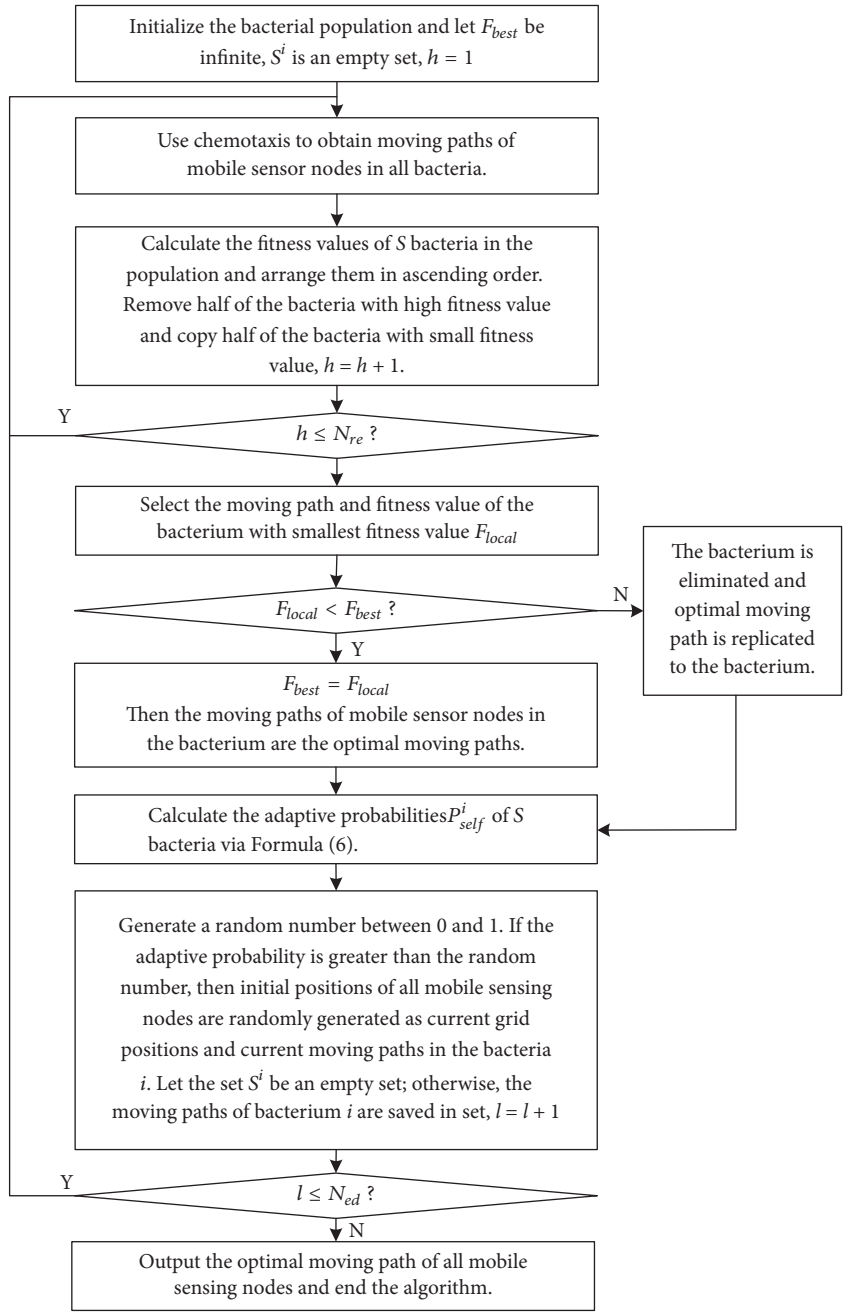


Figure 3: Flow chart of SCA_SM algorithm.

TCM follows the shortest moving path in the literature [10] which fully covers the monitoring area. As shown in Figure 7, no matter the number of mobile sensor nodes, SCA_SM can establish a path planning model of multisensor nodes' movement, calculate bacterial fitness value of each bacterium, solve the optimization model (1) by improving the bacterial foraging algorithm, and obtain optimal moving paths. Every mobile sensor node passes through several grids and moves to the vicinity of sink node to send data. Then SCA_SM reduces the packet loss rate and data transmission delay.

The bacterial population completes chemotaxis operation and replication operation of all mobile sensor nodes, then it implements and completes a migration operation. After each migration operation, mobile sensor nodes have their current optimal moving paths with different steps. The fitness value of optimal moving paths is current optimal fitness value. As shown in Figure 8, the abscissa represents the number of migrations, and the ordinate represents the optimal fitness value. The fitness value of SCA_SM decreases almost vertically after previous migration. After five migrations, the fitness value of optimal moving path drops to a lower value. As migration operation can help the algorithm jump out of local optimum and improve the convergence accuracy, the fitness value converges to 0.13424 after about 16 migrations.

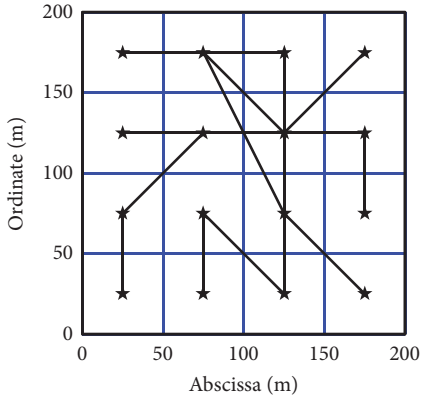


Figure 4: Moving paths of RAND_D.

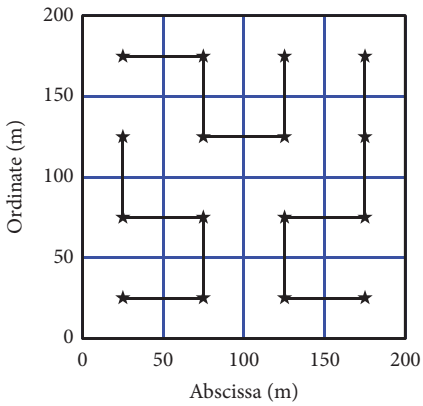


Figure 5: Moving path of HILBERT.

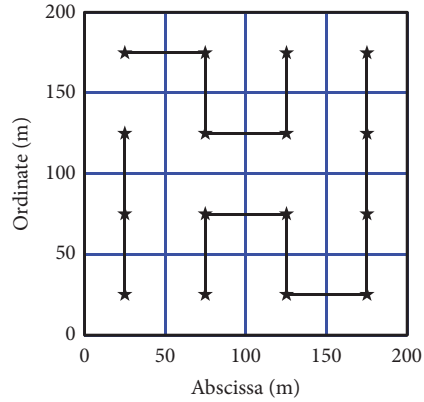


Figure 6: Moving paths of TCM.

This fitness value is the optimal solution of optimization model (1). Therefore, SCA_SM is convergent.

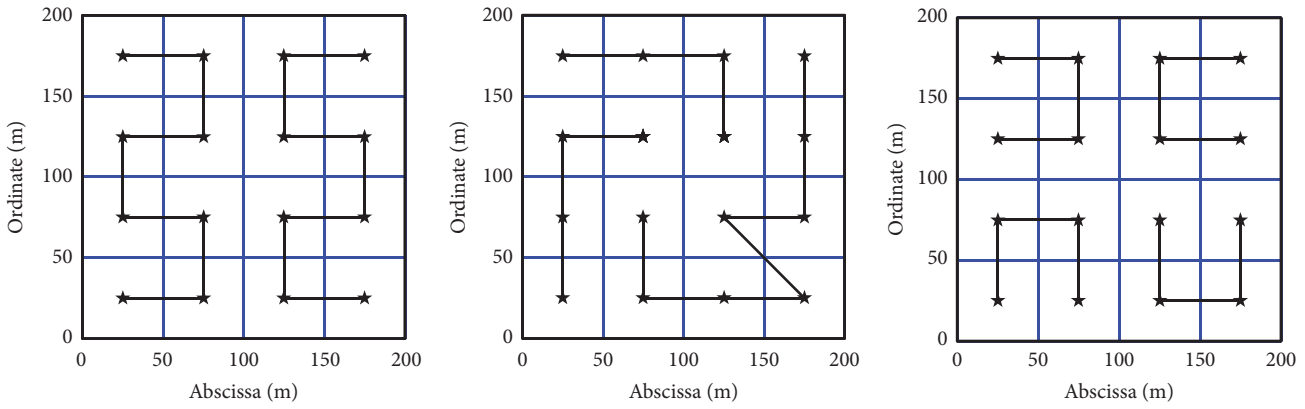
Simulation Results for Different Maximum Storage Spaces. In the section, the parameters in Section 5.1 are selected. Side length of the monitoring area is 400 m. Number of mobile sensor nodes is 3. Maximum storage space is 20, 22, 24, 26, 28, and 30 kbit. 10 different initial position distributions of the nodes are randomly generated. The packet loss rates and data transmission delays of RAND_D, HILBERT, TCM, and SCA_SM under each initial position distribution of nodes are calculated separately, and the average values are taken as simulation results.

As shown in Figure 9, in RAND, although maximum storage space of the node is increased, mobile sensor node randomly selects moving path. It leads that some moving paths are away from sink node and improves the packet loss rate. Therefore, packet loss rate of RAND fluctuates with the increase in the maximum storage space. Packet loss rates of HILBERT, TCM, and SCA_SM decrease with the increase in the maximum storage space. Especially, packet loss rate of SCA_SM decreases faster. Because HILBERT and TCM do not consider the data loss problem when storage space is full, packet loss rates of HILBERT and TCM are larger than that of SCA_SM. The SCA_SM uses bacterial foraging algorithm to solve the optimization model (1) and finds moving paths that weigh packet loss rate and data transmission delay through multiple migrations. In the moving path selection, when the storage space is almost full, the node moves to the communication range of sink node as much as possible, sends the stored data to sink node, and continues to perform mobile coverage. It can reduce the packet loss rate. In summary, when maximum storage space increases, packet loss rate of SCA_SM is always lower than that of RAND_D, HILBERT, and TCM. When maximum storage space reaches 34 kbit, the packet loss rate is close to 0%.

As shown in Figure 10, RAND_D has randomness, and data transmission delays of HILBERT and TCM decrease with the increase in the maximum storage space. However, they do not consider the data transmission delay in the moving path selection. SCA_SM establishes an optimization model, which is solved by improved bacterial foraging algorithm, and obtains the optimal moving path of each mobile sensor node which is through the grid centers in the communication range of sink node. When mobile sensor node moves along its optimal path, it can send its own data to sink node in time. Therefore, the data transmission delay of SCA_SM is always lower than the data transmission delays of RAND_D, HILBERT, and TCM.

Simulation Results in Different Monitoring Areas. Considering the size of the monitoring area, the number of square grids divided in the monitoring area is different. To analyze the impact of different sizes of monitoring area on performance parameters, number of mobile sensor nodes is 3, and maximum storage space is 1/4 of the number of grids. Square monitor area 1 has a side length of 800 m. Monitor area 2 has a side length of 700 m. Monitor area 3 has a side length of 500 m. Monitor area 4 has a side length of 400 m. Monitor area 5 has a side length of 200 m. The parameters in 4.1 are selected, and 10 different initial position distributions of nodes are randomly generated. The packet loss rate and data transmission delay of each algorithm are calculated under each initial position distribution of sensor nodes, and the average values are used as simulation results.

As shown in Figures 11 and 12, in the different monitoring areas, RAND_D has a large number of loss packets and



(a) Moving paths of two sensor nodes (b) Moving paths of three sensor nodes (c) Moving paths of four sensor nodes

Figure 7: Moving paths of SCA_SM.

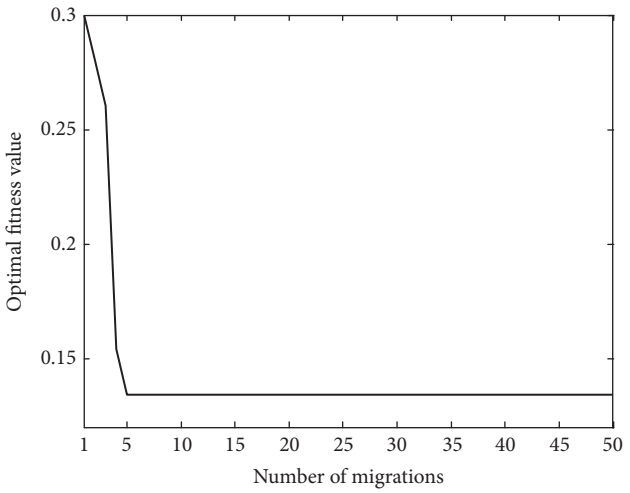


Figure 8: Convergence graph of SCA_SM.

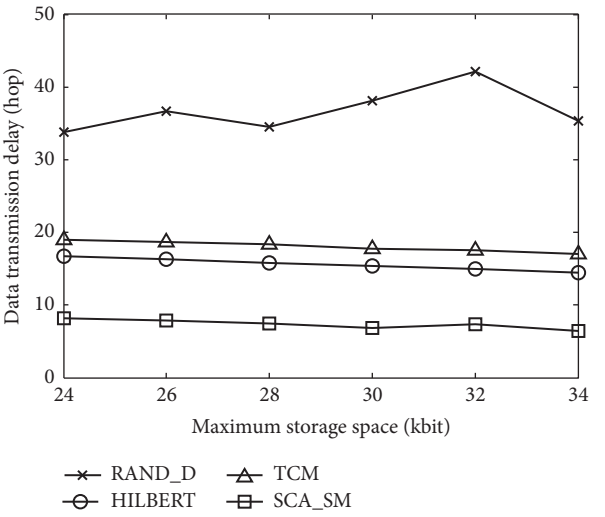


Figure 10: Comparison of data transmission delays for different maximum storage spaces.

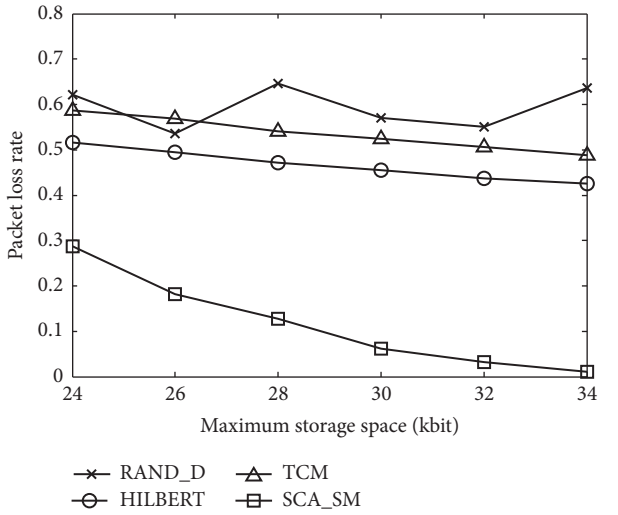


Figure 9: Comparison of packet loss rates for different maximum storage spaces.

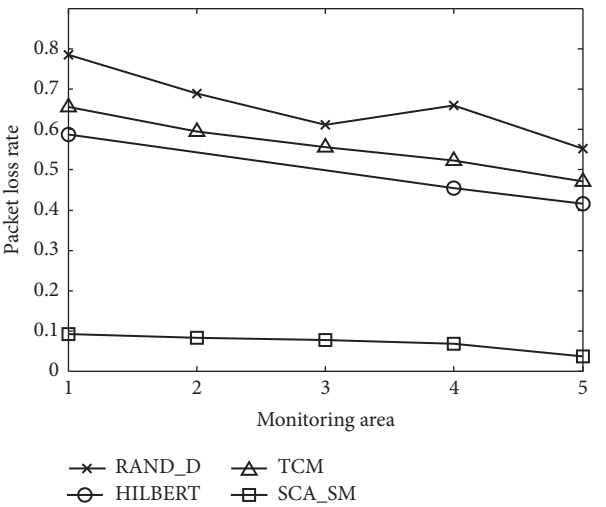


Figure 11: Comparison of packet loss rates in different monitoring areas.

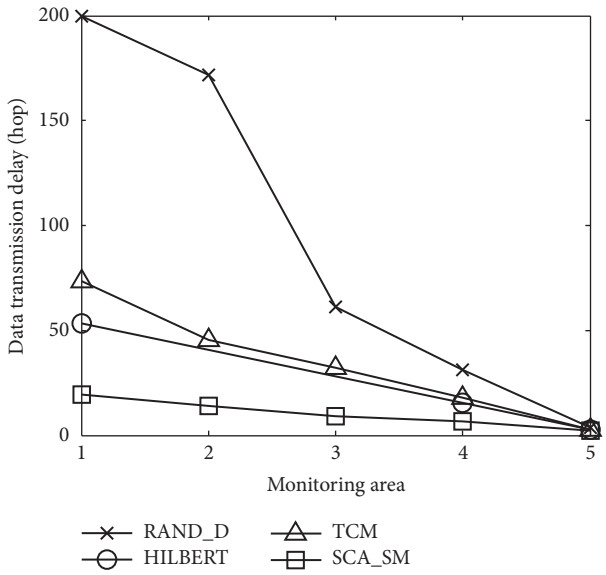


Figure 12: Comparison of data transmission delays in different monitoring areas.

a long moving path; therefore, its packet loss rate and data transmission delay are relatively large. The number of stayed grid of HILBERT must be a power of 4, so only the monitoring areas 1, 4, and 5 are valid and the moving path of each mobile sensor node is fixed. Its perceptual coverage problem of mobile sensor node is not considered. TCM only covers the entire monitoring area as short as possible and does not consider the packet loss rate and data transmission delay as HILBERT. SCA_SM establishes a reasonable optimization model and fitness function due to comprehensive consideration of packet loss rate and data transmission delay. Moreover, it uses improved bacterial foraging algorithm to obtain the optimal moving paths of all mobile sensor nodes. The moving optimal paths can maintain low packet loss rate and data transmission delay, and it is significantly better than RAND_D, HILBERT, and TCM.

6. Conclusion

A sensing coverage algorithm of sparse mobile sensor node with trade-off between packet loss rate and transmission delay (SCA_SM) is proposed in this paper. First, the algorithm assumptions and basic principles are proposed. Second, path selection constraint, data capacity update, packet loss rate, data transmission delay, and other related parameters and conditions are considered. An optimization model for fully covering the monitoring area and balancing the packet loss rate and data transmission delay is established. Bacterial fitness value calculation, bacterial chemotaxis operation, and improved bacterial foraging algorithm are proposed to solve the optimization model. Optimal moving paths of multiple mobile sensor nodes are obtained. Finally, simulation parameters of the algorithm are given. The simulation results show that the SCA_SM algorithm can find the optimal moving path of multiple mobile sensor nodes, has low data packet loss rate and data transmission delay, and is superior to RAND_D, HILBERT, and TCM. However, the algorithmic complexity of SCA_SM is relatively high. Therefore, the next goal is to study the optimization model and heuristic method of each mobile sensor node to solve the optimization model based on local information. Then optimization solution of each mobile sensor node is obtained, and it reduces the complexity of the algorithm.

Data Availability

The simulation data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 6 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Public Welfare Technology Application and Research Projects of Zhejiang Province of China under Grants no. LGF19F010006, no. LGG19F010011, no. LGF19F010005, and no. LGF18F010005, the National Natural Science Foundation of China under Grant no. 61501403, Education Department Project of Zhejiang Province of China under Grant no. Y201840757, and Science and Technology Project of Zhejiang Water Resources Department of China under Grant no. RC1850.

References

[1] W. H. Wang, T. Wang, Q. Wu et al., "Survey of delay-constrained data collection with mobile elements in WSNs," *Journal of Computer Research and Development*, vol. 54, no. 3, pp. 474–492, 2017.

[2] J. Habibi, H. Mahboubi, and A. G. Aghdam, "A gradient-based coverage optimization strategy for mobile sensor networks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 477–488, 2017.

[3] S. Rashed and M. Soyturk, "Effects of UAV mobility patterns on data collection in wireless sensor networks," in *Proceedings of the 4th IEEE International Conference on Communication, Networks and Satellite (COMNESTAT '15)*, pp. 74–79, IEEE, Indonesia, December 2015.

[4] Q. R. Zhang, R. X. Wei, R. K. He et al., "Path planning for unmanned aerial vehicle in urban space crowdedwith irregular obstacles," *Control Theory & Applications*, vol. 32, no. 10, pp. 1407–1413, 2015.

[5] J. R. Ding, C. P. Du, Y. Zhao et al., "Path planning algorithm for unmanned aerial vehicles based on improved artificial potential field," *Journal of Computer Applications*, vol. 36, no. 1, pp. 287–290, 2016.

[6] M. Zhao, Y. Yang, and C. Wang, "Mobile data gathering with load balanced clustering and dual data uploading in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 770–785, 2015.

[7] A. Kaswan, P. K. Jana, and M. Azharuddin, "A delay efficient path selection strategy for mobile sink in wireless sensor networks," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '17)*, pp. 168–173, IEEE, 2017.

[8] X. Li and Y. J. Wang, "WSN mobile aggregation node track design based on Hilbert space filling curve," *Modern Electronics Technique*, vol. 39, no. 23, pp. 17–21, 2016.

[9] L.-H. Hung, Y.-W. Huang, and C.-C. Lin, "Temporal coverage mechanism for distinct quality of monitoring in wireless mobile sensor networks," *Ad Hoc Networks*, vol. 21, no. 5, pp. 97–108, 2014.

[10] Y. Yue, J. Li, H. Fan, and Q. Qin, "Optimization-based artificial bee colony algorithm for data collection in large-scale mobile wireless sensor networks," *Journal of Sensors*, vol. 2016, Article ID 7057490, 12 pages, 2016.

[11] A. A. Ari, I. Damakoa, A. Gueroui et al., "Bacterial foraging optimization scheme for mobile sensing in wireless sensor networks," *International Journal of Wireless Information Networks*, vol. 24, no. 3, pp. 254–267, 2017.

[12] Z. D. Wang, E. L. Chen, and Z. D. Hu, "Sensor node deployment strategy of chaotic optimization of bacterial foraging algorithm," *Chinese Journal of Sensors and Actuators*, vol. 31, no. 1, pp. 110–118, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7810546, 17 pages
https://doi.org/10.1155/2019/7810546

Review Article

LMR and LTE for Public Safety in 700 MHz Spectrum

Aizaz U. Chaudhry and Roshdy H. M. Hafez

Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada K1S 5B6

Correspondence should be addressed to Aizaz U. Chaudhry; auhchaud@sce.carleton.ca

Received 7 March 2019; Revised 3 May 2019; Accepted 20 May 2019; Published 10 June 2019

Guest Editor: Maurizio Casoni

This paper presents a concise overview of current public safety communication networks known as LMR (Land Mobile Radio) and emerging LTE- (Long-Term Evolution-) based broadband public safety networks to be deployed in the 700 MHz band. A broadband nationwide network for public safety based on LTE is inevitable where shared or dedicated types of LTE-based public safety networks are possible. Current LTE services do not meet mission-critical requirements and several enhancements have been defined by 3GPP to address this in Releases 12 and 13. First responders are familiar with LMR and consider it to be a reliable technology with massive deployment everywhere. Therefore, it is expected that LMR will continue to exist alongside any new LTE-based broadband public safety network. Recent LTE releases (particularly Release 15) addressed the LMR-LTE interoperability issue and described comprehensive interworking facilities. New and upcoming features and services of LTE in Releases 14 and 15, such as mission-critical data, mission-critical video, and aerial user equipments, are also directly applicable to public safety. The paper endeavours to provide a quick yet meaningful review of all these issues. It also offers a look ahead at the new and rapidly advancing virtualization technologies, such as software-defined radio access network, and radio access network slicing, as enablers for future public safety networks.

1. Introduction

Public safety organizations have the responsibility to protect people's lives in natural and man-made disasters as well as in emergency situations. Such organizations include law enforcement agencies, emergency medical services, and fire departments; they are the first to arrive on the scene of emergency and are commonly referred to as *first responders*. The ability of first responders to flawlessly communicate among themselves and to access as well as share critical information in timely fashion influences their ability to save lives [1].

Public safety networks are perceived as mission-critical. They are required to be dependable, resilient, and secure, while satisfying other strict requirements concerning network coverage, system accessibility, and end-to-end performance. These crucial operational demands are the primary drivers for the design and engineering of the "*public safety grade*" network. Basic services provided by public safety networks include functionalities such as push-to-talk, group call, direct call, and dispatch services, which differentiate public safety services from typical user services over commercial networks [2].

The safety of public as well as the first responder personnel is based on the first responder's immediate access to reliable voice communications. *Push-to-Talk* (PTT) provides near instant call setup by allowing half-duplex communications between first responders, using a push button to instantaneously switch from voice reception mode to transmission mode. *Direct call* for off-network (or infrastructure-less) peer-to-peer communications, i.e., voice communications directly between two radio devices, must be provided in all operational environments. A *group call* entails the communication of speech to all group members. The permission to speak is administered by a dispatcher. *Dispatch services* consist of personnel with supervisory authority who manage and coordinate the activities of the first responders.

The mission-critical aspect of public safety communications places unique demands on the underlying radio technologies. When a first responder needs to use the PTT

service and presses the button on their radio device to request to talk, they must have the confidence that their voice communication will be successful with very high probability. In an emergency situation when the network may be congested, their call must go through anyway or lives may be lost.

Due to longer propagation distances, the 700 MHz band is an attractive option to build systems for public safety networks. For public safety, 20 MHz of dedicated spectrum is available for broadband while 12 MHz is available for narrowband communications in the 700 MHz band. *Land Mobile Radio* (LMR) communication systems for public safety are typically voice-centric narrowband systems. They operate in either the UHF, VHF, 700 MHz, or 800 MHz and 12.5 kHz is the standard bandwidth. *Long-Term Evolution* (LTE) is a widely deployed broadband technology offering high-data rate applications currently not supported in LMR.

Local public safety agencies in the U.S. have the flexibility to design and run their own public safety communication networks. Failures due to lack of interoperability between these disparate systems is considered a major problem for public safety. For example, interoperability issues on site after shooting at Columbine High School in 1999 forced first responders from different public safety agencies to use runners to carry written messages from one agency's command center to another, which greatly impacted their reaction time. Public safety networks also suffer from scalability issues such as holes in coverage, which affects their ability to always meet the mission-critical standards [3]. Hence arises the need for a single nationwide wireless communication network for all public safety agencies. Such a public safety network can be either based on sharing an existing nationwide commercial LTE network via upcoming technologies like radio access network slicing or can be a completely separate nationwide wireless network dedicated for public safety agencies such as FirstNet. FirstNet will be the first LTE-based dedicated nationwide public safety network in the U.S. that will operate in the 700 MHz spectrum.

As public safety embraces LTE, LMR features like direct call, group call, and PTT will need to be replicated onto LTE. Although LTE already provides services like PTT over cellular, over-the-top PTT, push-to-video, and push-to-x, these services do not meet the requirements of the mission-critical public safety radio users. To address public safety applications, 3GPP has defined several LTE enhancements in Releases 12 and 13 like proximity services, group communication system enablers, and mission-critical PTT. It may take several years for the transition from LMR to LTE. In the meantime, public safety agencies will most likely be using a mix of LMR and LTE networks, and effective interoperability solutions for these systems will be required.

3GPP has defined additional mission-critical services like mission-critical data and mission-critical video in Release 14. In addition to evaluating further mission-critical related topics like interworking between LTE and LMR for voice and data in Release 15, 3GPP is also investigating LTE's capability to provide connectivity to unmanned aerial vehicles as aerial user equipments (UEs), which could be vital to using unmanned aerial vehicles for public safety operations. Unmanned aerial vehicles are also being considered for deployment as aerial base stations in LTE-based networks to restore critical communications during disasters. Enabling technologies like software-defined networking, network functions virtualization, software-defined radio access network, and radio access network slicing will also play an important role in shaping future public safety networks. *The contribution of this work includes presenting a brief overview of 700 MHz spectrum, LMR, and LTE in relation to public safety communications; exploring shared and dedicated type of LTE-based public safety networks; highlighting existing non-mission-critical public safety services over LTE as well as mission-critical enhancements to LTE such as device-to-device communications (or proximity services), group communication system enablers, and mission-critical PTT; discussing coexistence of LMR and LTE and examining solutions for their interoperation; highlighting new and upcoming features and services of LTE directly applicable to public safety such as mission-critical data, mission-critical video, and unmanned aerial vehicles as aerial UEs; and providing a look ahead at the emerging virtualization technologies, such as software-defined radio access network, and radio access network slicing, as enablers for future public safety networks.*

An overview of legacy and emerging public safety communication technologies is presented in [1]. Introduction of novel capabilities in LTE for public safety has been discussed in [2]. Policies leading to present day public safety communication systems, alternative directions for the future, and steps toward a more effective policy are highlighted in [3]. Suitability of LTE for mobile broadband public safety services is examined in [4]. Public safety use cases are described in [5] along with current status of related activities in 3GPP standards, and future challenges in public safety. Legacy public safety networks and their limitations, potential of LTE for future public safety networks, and rapid emergency deployment in LTE-based public safety networks are surveyed in [6] along with some future research challenges.

A high-level discussion of LMR, LTE, and Voice over LTE (VoLTE) is provided in [7] along with recommendations for possible FirstNet architectures. Operational contexts and requirements of public safety organizations, different wireless communication technologies used by public safety organizations, technology standards and regulatory frameworks governing public safety organizations, and potential evolution of communication technologies in the public safety domain are discussed in [8]. Some aspects related to public safety communications that are reviewed in this paper that have not been explored in [1–8] include existing non-mission-critical public safety services over LTE including over-the-top push-to-talk, push-to-video, and push-to-x; LMR-LTE interoperability solutions for public safety including interworking function, inter-RF subsystem interface, and radio over IP; current research issues including mission-critical data, mission-critical video, and use of unmanned aerial vehicles as aerial LTE UEs for public safety operations; and enabling technologies for future public safety networks including software-defined radio access network, and radio access network slicing.

The rest of the paper is organized as follows. A brief overview of 700 MHz radio spectrum, LMR, and LTE in

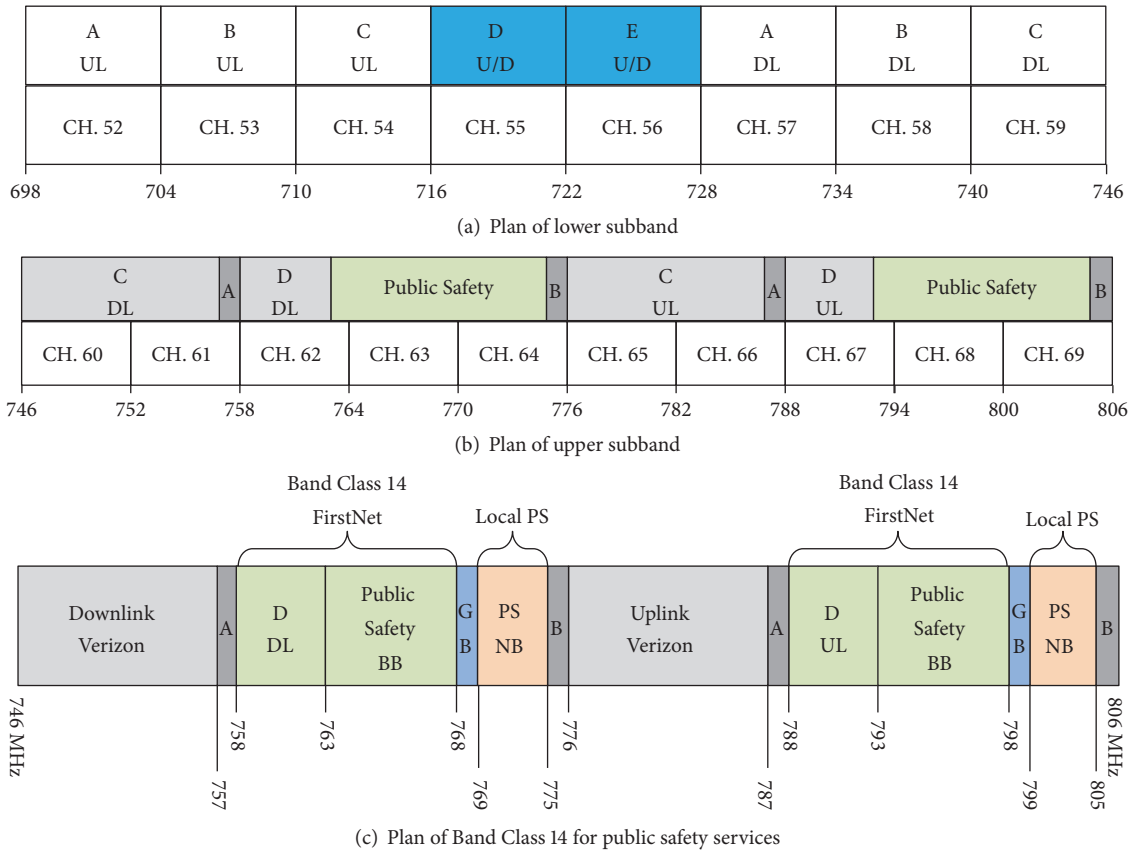


Figure 1: 700 MHz spectrum.

perspective of public safety communications is provided in Sections 2, 3, and 4, respectively. Section 5 discusses shared and dedicated type of LTE-based networks for public safety. The existing non-mission-critical public safety services over LTE are briefly described in Section 6. Section 7 gives details of LTE enhancements for mission-critical services while LMR-LTE interoperability solutions are discussed in Section 8. Section 9 presents current research issues and Section 10 highlights enabling technologies for future public safety networks. The conclusions are summarized in Section 11.

2. 700 MHz Spectrum for Public Safety

Wireless signals in the 700 MHz band have good propagation characteristics, which makes this band an attractive option to build wireless systems for commercial as well as public safety networks. The 700 MHz band is made up of two subbands: the lower 700 MHz subband and the upper 700 MHz subband. Figures 1(a) and 1(b) illustrate the spectrum allocation, uplink (UL) and downlink (DL) frequency, and band gap for these subbands. In Band Class 14 in the U.S., 20 MHz of dedicated spectrum in the 700 MHz band is allocated for public safety as illustrated in Figure 1(c). This broadband (BB) spectrum, highlighted using green colored blocks, is intended for an LTE-based public safety network. The D block will be reassigned for utilization by Public Safety BB. Consequently, 10 MHz will be available for downlink and 10 MHz will be available for uplink for Public Safety BB in Band Class 14. Narrowband (NB) spectrum, shown in orange colored blocks, is designated for narrowband public safety networks, such as LMR systems for voice communication. To prevent any interference, guard bands (GBs) of 1 MHz are positioned between narrowband, broadband, and commercial spectrum [1].

3. LMR for Public Safety

LMR is a wireless communication system meant for terrestrial users operating portable or mobile radio units like two-way digital radios or walkie-talkies. Public safety LMR systems are typically voice-centric and are purposely built to support individual and group communications. In a traditional LMR network configuration, shown in Figure 2, the first responders are connected to dispatch services and to each other via trunked radio access points or repeaters.

LMR started as a mechanism for sending analog voice notifications from an AM base station to a mobile radio receiver and quickly transformed into a two-way FM system. During the next major development that came under the Association of Public-Safety Communications Officials (APCO) International's Project 25 (P25) [9], LMR shifted to digital (or analog) radios, which could operate in 12.5 kHz channels on either the VHF, UHF, 700 MHz, or 800 MHz

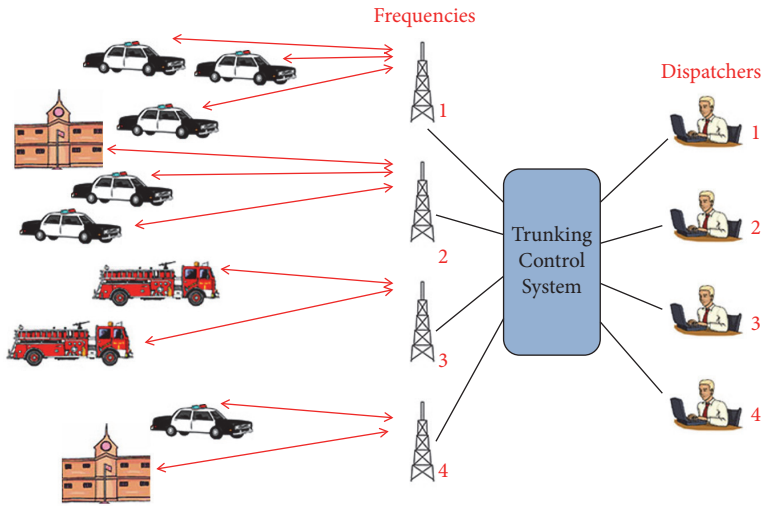
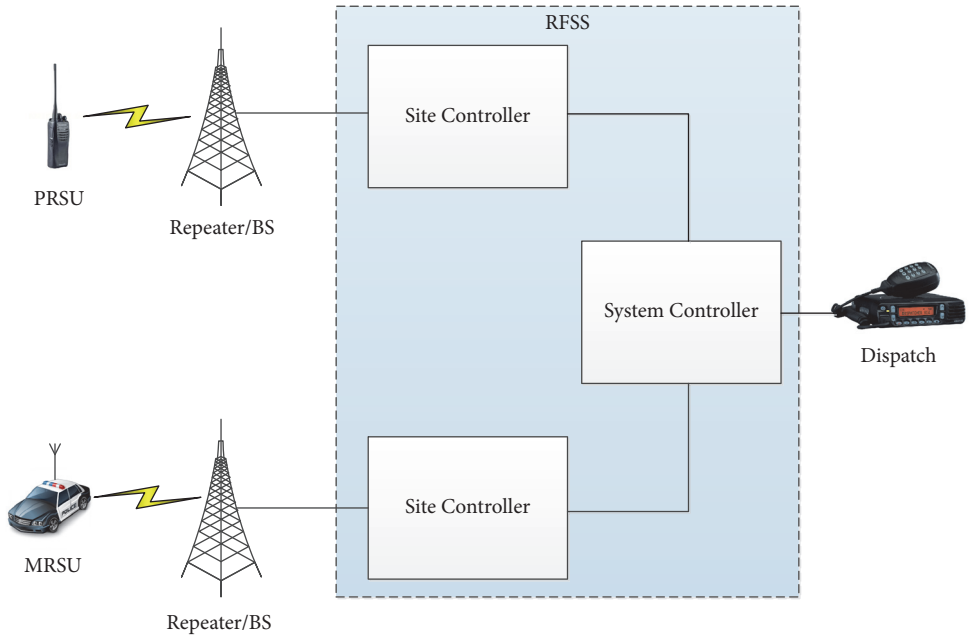


Figure 2: A traditional LMR network.



RFSS – Radio Frequency Subsystem
PRSU – Portable Radio Subscriber Unit
BS – Base Station
MRSU – Mobile Radio Subscriber Unit

Figure 3: System-level representation of LMR network.

bands. P25 forms the core function of majority of LMR radio communication systems for public safety [7].

P25 radios use analog mode to communicate with legacy radios, and either digital or analog mode for communication with other P25 radios. P25 supports voice and low-data rate communications limited to a maximum of 9600 bits/s. It does not subscribe to any cellular radio structure; instead, the coverage is based on strong wide-coverage transmitters with extensions and repeaters. The usual coverage radius of a base station is a few kilometers depending on the terrain. P25 offers a prolific set of features including PTT, direct call, group call, and others. At the operational level, features offered by P25 exhibit real demands of public safety operations since it was designed with end users' needs as requirements [8].

P25 radio systems are either conventional or trunked. A *conventional* P25 system consists of a simple infrastructure of repeaters that repeats radio calls from one frequency channel to another. A *trunked* P25 system is characterized by a central controller in the infrastructure, which assigns calls to available channels. Instead of giving each user a dedicated channel, channels are assigned to users as needed. System-level representation of LMR system is shown in Figure 3.

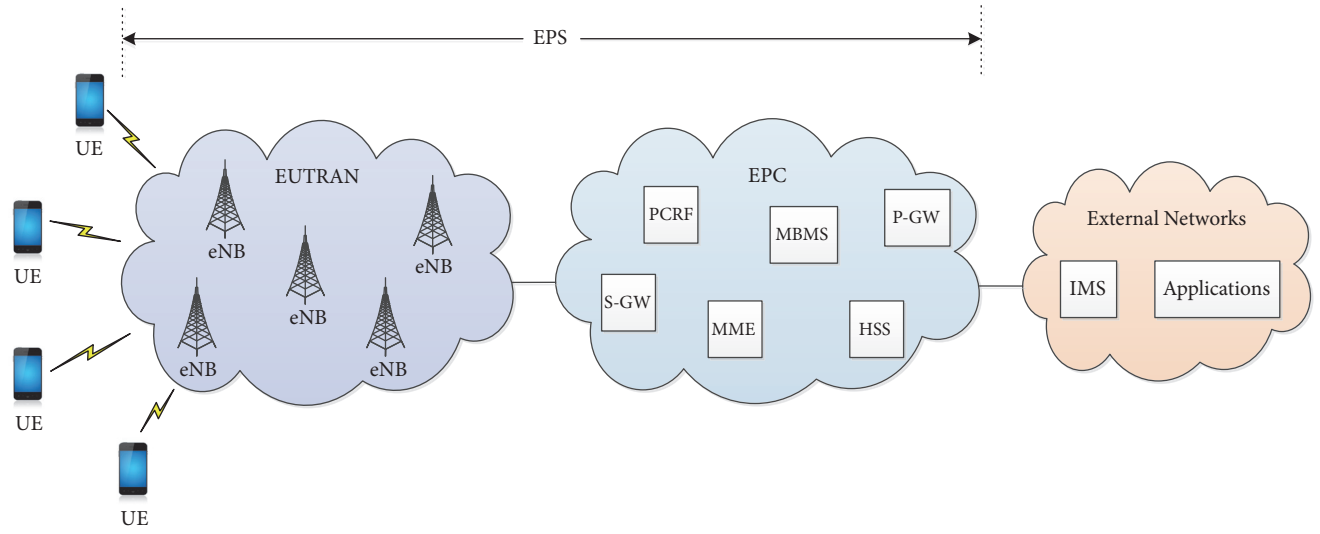


Figure 4: System-level representation of LTE network.

The LMR system consists of mobile or portable subscriber units, and repeaters or base stations connected to a Radio Frequency Subsystem (RFSS).

Terrestrial Trunked Radio (TETRA) [10] is a well-established narrowband technology that delivers mission-critical communications to first responders in Europe. The TETRA technology has been standardized by the European Telecommunication Standards Institute keeping in view the requirements of the public safety agencies. TETRA for Police (TETRAPOL) [11] is another technology for public safety communications that competes with TETRA to provide narrowband public safety networks in Europe. Other professional mobile radio technologies including TETRA and TETRAPOL are not discussed here in detail since the focus of this paper is on LMR.

4. LTE for Public Safety

LTE has been broadly deployed as the worldwide mobile broadband standard. The primary benefit in using LTE for public safety is to take advantage of all technological advances being constantly introduced into commercial cellular systems. LTE is a broadband technology that allows high-data rate applications currently not supported in LMR. LTE offers an all-IP system architecture and flexible air interface that supports carrier bandwidths from 1.4 MHz to 20 MHz [2]. LTE is currently able to transmit data at 300 Mbps on the downlink and 75 Mbps on the uplink. It is expected that the data rate of LTE-based networks will increase substantially over the next few years.

Figure 4 illustrates the system-level block representation of LTE. An LTE network consists of two key parts: the Evolved Universal Mobile Telecommunications System (UMTS) Terrestrial Radio Access Network (E-UTRAN) [12] and the Evolved Packet Core (EPC) [13]. Together, E-UTRAN and EPC comprise the Evolved Packet System (EPS). EPS employs EPS bearers to provide IP connectivity to a UE for accessing the Internet. E-UTRAN controls the radio transmission functions, whereas EPC handles the session and mobility management functions.

E-UTRAN comprises base stations named evolved NodeBs (eNBs or eNodeBs). EPC consists of a Mobility Management Entity (MME) for control functions including location management, a Serving Gateway (S-GW) for managing user traffic from/to the E-UTRAN to/from the EPC, and a Packet Data Network Gateway (P-GW) for providing IP connectivity to external IP networks by setting up EPS bearers between the UE and P-GW. Home Subscriber Server (HSS) is a central database that contains information related to users' subscription. IP Multimedia Subsystem (IMS)—an IP-based service control platform—is used to support advanced multimedia services. The Policy and Charging Rules Function (PCRF) entity controls the treatment and charging functionalities of different data flows over EPS bearers.

Multimedia Broadcast Multicast Service (MBMS) is a point-to-multipoint service for transmitting data from a single source to multiple receivers. MBMS was introduced in Release 6 and has been updated in Release 9 [14], where it is known as *evolved MBMS* (eMBMS). The main constituents of eMBMS include Broadcast/Multicast Service Center (BM-SC), MBMS Gateway (MBMS-GW), MME, Multicell/Multicast Coordinating Entity (MCE), and eNB. BM-SC controls broadcasting services to end users and serves as point of entry to the mobile network for content providers

transmitting content from external networks. MBMS-GW broadcasts to eNB within the service area and eNB distributes the data to end users. MCE manages admission control and allocates radio resources for the MBMS session. eMBMS is being exploited by 3GPP for public safety operations over LTE and will play a vital role in providing LMR-like group communication services over LTE.

5. LTE-Based Networks for Public Safety

In LMR, first responders communicate over infrastructure and spectrum that is dedicated to public safety. However, these public safety networks do not always meet the mission-critical standards due to coverage holes. Many first responders like police and firefighters routinely carry cell phones as backup in case their public safety network fails to provide coverage. Also, public safety agencies in the U.S. design and run their own public safety communication networks, which leads to interoperability failures between these disparate systems [3].

To overcome these coverage and interoperability issues, a single nationwide wireless network is required for all public safety agencies. One option is to share an existing nationwide commercial LTE network either by sharing infrastructure or by prioritizing public safety traffic to guarantee capacity for public safety in case of an emergency. Another option will be to build a dedicated nationwide LTE-based wireless network that will be exclusive for public safety agencies. Yet another option can be to create a new nationwide wireless network to serve both public safety and commercial subscribers. The business case of such a network is discussed in [15].

5.1. Sharing LTE for Public Safety. An LTE-based public safety network can share a nationwide commercial LTE network via passive or active infrastructure sharing. Passive sharing, such as sharing sites or masts, does not need active coordination between different network operators. Active sharing involves sharing of active elements or intelligence of the network, such as RAN (Radio Access Network) or EPC sharing.

All 3GPP LTE releases support active infrastructure sharing, generally referred to as network sharing. The sharing scenarios can be divided into three main categories [16]:

(i) Geographically split network sharing: In this case, each operator has its EPC and RAN with no overlapping coverage. The networks of both operators (operator A and operator B) are interconnected, i.e., RAN-A is connected to EPC-A and EPC-B and similarly RAN-B is connected to EPC-A and EPC-B.

(ii) Multiple EPCs sharing a single RAN: In this case, each operator has its own EPC but these operators share a single RAN. A single RAN can be shared by multiple EPCs using RAN slicing. This is an upcoming technology that can be used for enabling future public safety networks and is discussed later in this paper.

(iii) Multiple RANs sharing a common EPC: In this case, each operator has its own RAN but these operators share a common EPC.

5.2. Dedicated LTE-Based Network for Public Safety. FirstNet is an LTE-based wireless network that gives first responders their own separate nationwide broadband network. The First Responder Network Authority, created by the U.S. Congress in February 2012, is charged with overseeing the construction, operation, and maintenance of FirstNet—the country's first nationwide public safety broadband network. In the spring of 2017, a giant step forward for FirstNet was the start of a public-private partnership between AT&T and First Responder Network Authority to build a $46.5 billion broadband network. AT&T has been tasked to build, deploy, operate, and maintain this network under a 25-year agreement. After roughly a year of partnership, FirstNet is starting to come alive as AT&T announces the nationwide launch of its dedicated network core in March 2018. FirstNet's own EPC built on dedicated (or physically separate) hardware implies that first responders have a separate broadband wireless network and that its traffic is isolated end-to-end, as the FirstNet core totally segregates public safety traffic from all commercial traffic.

FirstNet is an LTE-based wireless broadband network exclusive for public safety services that operates in Band Class 14 of the 700 MHz spectrum. Salient features of FirstNet comprise PTT, direct communication mode, group calls, full duplex voice system, talker identification, and emergency alerting. As shown in Figure 5, FirstNet comprises distributed core, terrestrial mobile systems, satellite mobile systems, and deployable mobile systems. Distributed core incorporates an EPC network and a service delivery platform to dispense different services to end users. The terrestrial mobile systems consist of terrestrial-based communication, while satellite mobile systems will employ satellite communication links to connect to the satellite core network. Deployable mobile systems are mainly comprised of cells on wheels (or vehicles), which provide services in areas with network congestion or fill coverage holes [1].

FirstNet vehicular networks are split into five categories: vehicle network system (VNS), cell on light truck (COLT), cell on wheels (COW), system on wheels (SOW), and deployable aerial communications architecture (DACA). These FirstNet vehicular systems are envisioned to serve a vital part in arranging coverage extension for the nationwide public safety broadband network. Such deployments will provide necessary coverage and capacity in areas without terrestrial coverage, or where normal coverage disappears during natural or man-made disasters.

During disasters, such as earthquakes, critical communication facilities are likely to be without power. After a few hours, widespread communication outages can be expected as backup batteries and generators at these facilities begin to fail. Terrestrial communication facilities can also be damaged resulting in significant communication outages. DACA has the potential to dramatically improve emergency response in such disaster situations. It is envisioned as an aerial capability that is deployable within the first few hours after a disaster to temporarily restore critical communications. In the presence of DACA systems, first responders on the ground can continue to provide emergency services while repair crews restore the terrestrial communication infrastructure [17].

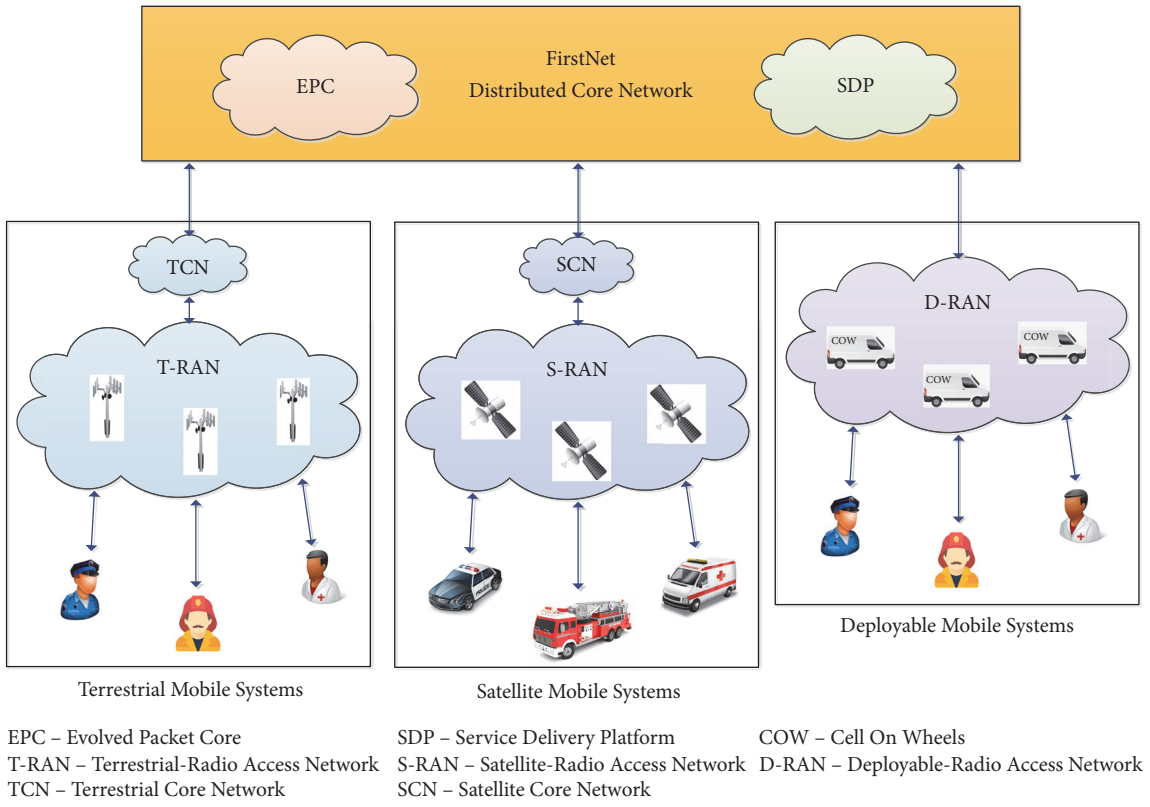


Figure 5: FirstNet's architecture for public safety communications.

6. Existing Non-Mission-Critical Public Safety Services over LTE

Non-mission-critical services over cellular (or LTE) provide users with the possibility to create or access talkgroups and set up voice, data, or video calls within these talkgroups by pushing a key on the handset. However, these cellular services do not satisfy the demands of the mission-critical public safety radio users in areas of guaranteed and seamless network coverage and capacity in rural and urban localities during incidents.

6.1. Push-to-Talk over Cellular (PoC). PTT services over cellular have been available for many years with commercial network solutions. The concept of carrier deployed PTT or *Push-to-Talk over Cellular* (PoC) or cellular PTT has continued with Sprint in their QChat-enabled service and Kodiak-enabled services for Verizon and AT&T [18]. Verizon offers a PoC service on their network called PTT+. This service is provided via a hosted solution from Kodiak Networks.

PoC is a mobile telephony service for one-to-one and one-to-many half-duplex communications over cellular phones emulating two-way push-to-talk LMR radios. A single person can speak at a time and the rest of the participants hear this verbal communication. Other participants can reply to this speech once it has ended. *Floor control* handles the contention of the *right-to-speak* among various participants. Recipients hear the sender's voice either via *auto answer* in which case they do not require any effort on their side or through *manual answer* after being prompted to accept the connection.

Initial PoC systems were launched in the early 2000s to operate over 2G or 2.5G cellular networks. Cellular networks have quickly progressed to LTE, which supports mobile data applications requiring high data rates. To deliver high-quality one-to-one and one-to-many voice communications, the latest broadband PoC technology leverages LTE broadband cellular networks. Broadband PoC can operate on a wide range of cellular devices including ordinary smartphones or special rugged smartphones having a dedicated PTT button. Organizations that use LMR can employ broadband PoC for LMR augmentation [19].

6.2. Over-the-Top Push-to-Talk (OTT PTT). In *Over-the-Top* (OTT) PTT architecture, the PTT system is independent of the carrier's wireless network. OTT PTT applications work on a variety of devices and across multiple carrier networks as these are application-layer-enabled systems. These are typically cloud-based solutions that can be implemented immediately and are cost competitive and scalable [18]. Google Play has more than 150 applications available for download that offer PTT functionality. The main advantage of OTT PTT applications is their ability to work on multiple access technologies such as Wi-Fi and LTE.

OTT PTT solutions overlay the cellular network and are not integrated or optimized. Since these solutions utilize the Internet as a means to access the cellular network like any

other third-party data application, they are not subject to any Quality of Service criteria. Carrier deployed PoC, on the other hand, provides tight integration between the wireless network and PoC system, resulting in faster call setup times and better performance during periods of high network congestion.

6.3. Push-to-Video (PTV). Video utilises and amplifies a human's capability to collect a vast amount of visual information. A few seconds of video can provide much more data regarding an incident and the possible outcomes of a situation that cannot be described in voice communications or shared with still images. For first responders at the frontline, *Push-to-Video* (PTV) communications will mean improved situational awareness leading to improved first responder safety, which will in turn result in greater public safety.

The Group Communications solution from Nokia provides a push-to-video feature over cellular (or LTE) to enhance situational awareness. It supports the ability for groups of first responders to share live video feeds from a disaster scene.

High bandwidth, low latency, dedicated throughput, unlimited availability, and geographic coverage are the key requirements for effective use of PTV communications over LTE. A dedicated public safety LTE-based mobile broadband network such as FirstNet has great potential in this area.

6.4. Push-to-X (PTX). *Push-to-X* (PTX) is an evolution of PoC. However, it is more than just "pushing" voice to a user via PTT; it leverages the same application to push data like images or to push video to another user or groups of users. For instance, a first responder may wish to transmit information, other than speech, for example, data, text messages, GPS coordinates, a map, a video file, live audio, streaming audio, live video, streaming video, etc., to the talkgroup or to a colleague with the push of a button.

AT&T has recently upgraded their Enhanced PTT (EPTT) service. The upgrade features PTX functionality that enables sending highly secure texts, photos, videos, voice recordings, files, and location data to groups on FirstNet and AT&T network.

Today's public safety mobile devices for LMR are restricted to voice communication and are not capable of transmitting multimedia data to first responders. Expway and Bittium are offering LTE broadcast-enabled mobile devices, which support push-to-talk, high-definition push-to-video, and large file delivery for public safety communications.

7. LTE Enhancements for Mission-Critical Public Safety Services

As public safety adopts broadband technologies such as LTE, all existing features and applications need to be replicated onto that broadband platform, while preserving interoperability with existing narrowband public safety networks such as LMR. LMR employs a number of features like direct call, group call, and PTT, which were not taken into consideration when LTE was designed. Next-generation public safety communications in the U.S. can be envisaged as a ubiquitous 700 MHz LMR narrowband network overlaid with a 700 MHz LTE broadband network. LMR is expected to persist as the lifeline for public safety agencies. LTE will provide them access to high-data rate applications that cannot be sustained over narrowband wireless technologies like LMR.

3GPP concentrated on two major areas in LTE Release 12 to address public safety applications: *Proximity Services* (ProSe) [20] and *Group Communication System Enablers for LTE* (GCSE_LTE) [21]. An objective of 3GPP was to safeguard the quality of LTE while incorporating these features.

7.1. Proximity Services (ProSe). When the network coverage is absent, a public safety UE can automatically use proximity services. Even in the presence of network coverage, the UE can be manually set by the user to employ direct communication. In *conventional LTE communication*, the data path of two UEs, communicating in close proximity with each other, goes through the operator network where eNBs and/or gateways are involved. In *network-assisted direct communication* between UEs in close physical proximity that are served by the same eNB, the data path is routed locally for their communication. *Direct communication* between two UEs without network supervision moves the data path off the access and core networks onto direct links between the UEs. In direct communication, a communication link is established between the two users without traversing the network; this saves network resources while enabling mission-critical communication among the first responders even when they lack network coverage [1]. Direct communication does not need any support from the network, and communication is carried out by utilizing information locally available at the UEs. Network assisted direct communication needs network assistance for authorizing the connection.

The ProSe-based direct communication between ProSe-enabled UEs has been standardized in Release 13 in 3GPP's technical specification TS 23.303. The 3GPP system enablers for ProSe encompass the following functions [22]:

(i) EPC-level ProSe discovery

(ii) EPC support for WLAN direct discovery and communication

(iii) Direct discovery

(iv) Direct communication

(v) UE-to-Network relay.

7.2. Group Communication System Enablers for LTE (GCSE_LTE). A *group communication service* is meant for providing an efficient method to convey the same message to multiple users in a controlled fashion. Group communications is extensively used in LMR systems for public safety operations. The basic purpose of this service in LMR is to deliver PTT functionality; so a group communication service based on 3GPP architecture that uses LTE should support PTT voice communications with similar performance [23].

3GPP's technical specification TS 23.468 [24] describes how a Group Communication Service Application Server (GCS AS) may use the 3GPP system enablers to provide a

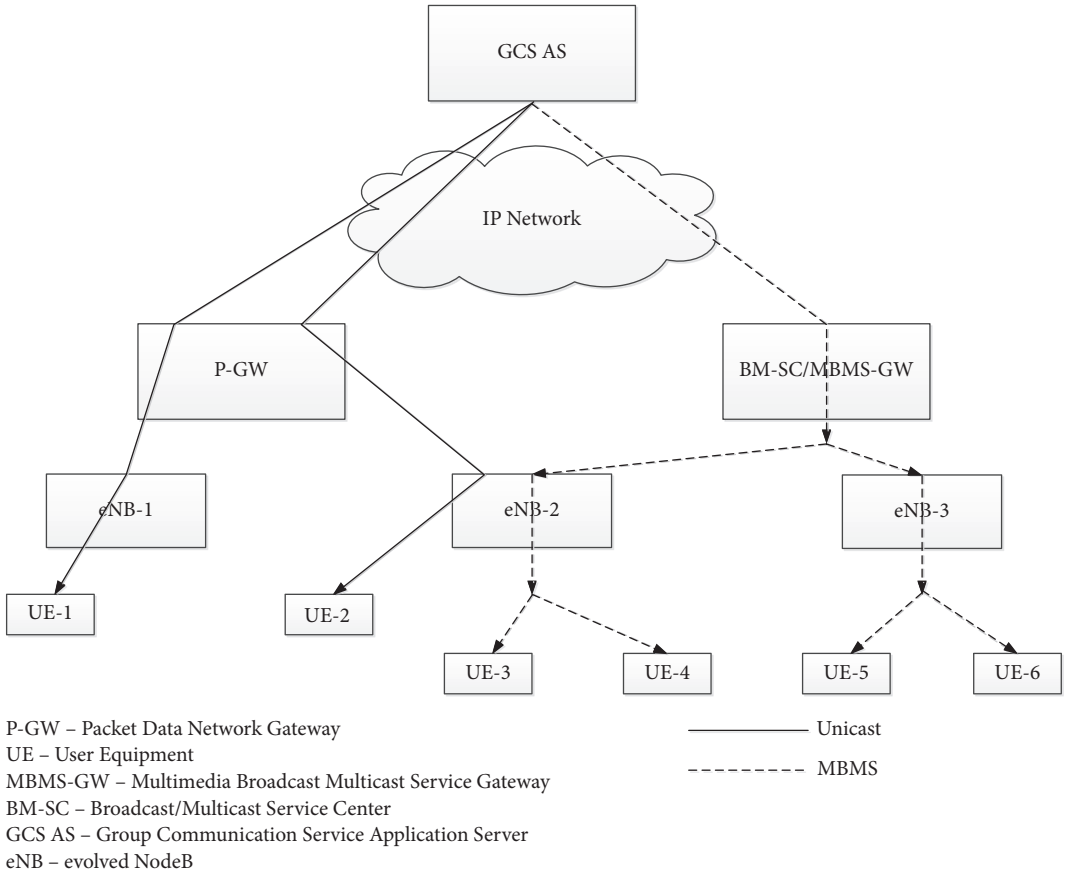


Figure 6: Combination of unicast and MBMS delivery for group communications.

group communication service. These enablers are known as *group communication system enablers*. The GCS AS employs EPS bearer services and may also employ MBMS bearer services. The UE uses an EPS bearer service to send data to the GCS AS in uplink. In downlink, the GCS AS may use the UE's individual EPS bearer service and/or MBMS bearer service to transfer data. Figure 6 shows a situation where the GCS AS employs a combination of unicast and MBMS delivery for different UEs belonging to a single group. UE-1 and UE-2 receive DL traffic through unicast whereas UEs 3–6 receive DL traffic via MBMS. Different delivery modes are used for UE-2, UE-3, and UE-4, even though they are connected to the same eNB (i.e., eNB-2). Unicast is used for UE-2 since it is in an area having low MBMS signal strength.

7.3. Mission-Critical Push-to-Talk (MCPTT). The *Mission-Critical Push-to-Talk* (MCPTT) service offers an enhanced PTT service over LTE that is appropriate for mission-critical scenarios. It emulates the behavior of Push-to-Talk service delivered by LMR. MCPTT is meant for providing voice communication between several users during a group call where each user can attain access to the permission to talk in a controlled fashion. However, it also offers direct (or private) calls between pairs of users [25].

First responders operate in groups and perform different tasks. Their tasks and operations are controlled, assisted, or coordinated by a dispatcher. For their communications, first responders are organized in groups. To enable them to coordinate quickly, people working together converse in the same MCPTT group. People with different functions converse in separate MCPTT groups. The routine public safety tasks are handled by standard procedures using dedicated MCPTT groups. However, MCPTT groups are also created for tackling large incidents.

3GPP has been diligently working to develop technical enhancements for LTE that support MCPTT. MCPTT builds upon service enablers: Group Communication System Enablers for LTE (GCSE_LTE) and Proximity Services (ProSe). The end user is anticipated to have similar experience whether MCPTT is employed under coverage of an EPC network or in direct communication when there is no network coverage. FirstNet is promising a mission-critical voice service—a VoLTE service with enhanced priority—in future in its roadmap. However, this is not MCPTT. Although 3GPP has completed its standardization of MCPTT in technical specification TS 22.179, implementation by manufacturers and service providers, such as FirstNet, has still to occur. LTE off-network communications is still not mature enough to match LMR direct call (or talk-around) service.

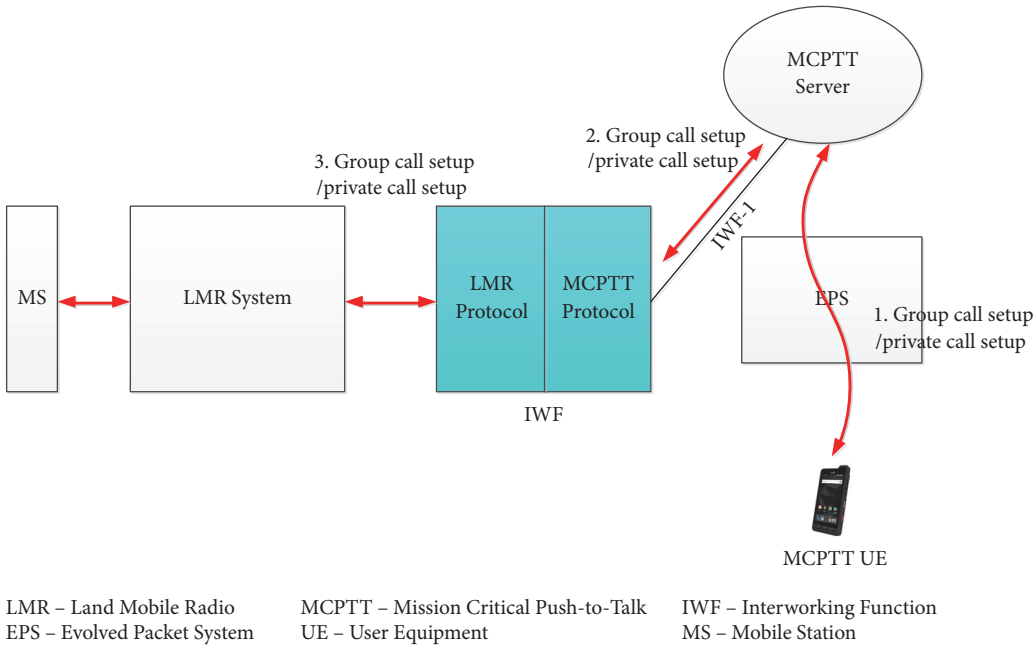


Figure 7: General interworking between LTE and LMR.

7.4. Open Mobile Alliance-Push-to-Talk over Cellular (OMA-PoC). Open Mobile Alliance (OMA) develops specifications for the application layer called service enablers. OMA enablers provide a standardized approach to tasks such as data gathering and transporting information from a network to a device or server. They are network agnostic, i.e., they are devised to work over any kind of network layer.

OMA's *Push to Communicate for Public Safety* (PCPS) is a PTT specification for LTE [26]. It has been adopted by 3GPP for use in defining the MCPTT requirements. PCPS has evolved from the *OMA-Push-to-Talk over Cellular* (OMA-PoC) enabler [27]. OMA-PoC employs Real-time Transport Control Protocol (RTCP) to carry PTT control information, Real-time Transport Protocol (RTP) for voice packets, and Session Initiation Protocol (SIP) as the call control protocol.

8. LMR-LTE Interoperability for Public Safety

LMR handsets usually transmit using 3 to 5 Watts of power, whereas an LTE handset may transmit at about 1 Watt. This simply results in longer range for LMR systems. For an LTE network to support the same coverage area as an LMR network, more sites spaced closer together need to be installed that will result in higher costs. Due to prohibitive infrastructure costs, it will not be easy for an LTE broadband network at 700 MHz to replace LMR.

LTE will not replace LMR any time soon. In the initial version of FirstNet, LTE will be a complimentary enabler that will sit on top of LMR. It may be several years before the transition from LMR to LTE is made. It is highly likely that public safety agencies will be using a mix of LMR and LTE networks in both the short and long term and will need to have effective interoperability solutions. The ultimate goal will be to achieve standards-based interworking between legacy PTT on LMR and MCPTT on LTE.

A report published in January 2018 by the LMR-LTE Integration and Interoperability Working Group of National Public Safety Telecommunications Council (NPSTC) has identified the following requirements for LMR-LTE interoperability [28]:

(i) First responders operating on LMR and LTE networks shall be able to communicate with each other

(ii) Consoles operating on LMR and LTE networks shall be able to monitor and participate in the voice communications on interworked LMR and LTE talkgroups

(iii) First responders and consoles shall have access to multiple LTE talkgroups to coordinate operations and many of these LTE talkgroups will need to be interconnected with LMR talkgroups to support these operations.

8.1. Interworking Function (IWF). 3GPP has studied and identified solutions in their technical report TR 23.782 [29] suitable for interworking between LTE mission-critical systems and non-LTE mission-critical systems that satisfy the MCPTT requirements. LMR system specifications define the equipment and subsystems that constitute the network including base stations and terminals, whereas, in LTE, the MCPTT server delivers centralized support for MCPTT services. In order to realize communication between these different systems, an *Interworking Function* (IWF) is introduced to support protocol translation, identity mapping, routing, and so on. Figure 7 illustrates a solution based on IWF. It is assumed that the interworking group, consisting of

group members from the MCPTT/LTE system and the LMR system, has been created and configured before a group call is initiated on this interworking group. Following are some of the interworking scenarios between LTE and LMR systems that have been discussed in this report along with related issues and their solutions:

(i) LTE UE initiates private communication to LMR Mobile Station (MS)

(ii) LMR MS initiates a private communication to LTE UE

(iii) LTE UE or LMR MS initiates a group communication.

8.2. Inter-RF Subsystem Interface (ISSI). The TIA-102.BACA-A [30] specification defines how RFSS can be connected via an IP interface to allow wireline interoperability. The *Inter-RF Subsystem Interface* (ISSI) is an IP based connection that uses SIP for call control and media is handled via RTP. ISSI specifications were defined to achieve network interoperability by interconnecting RFSSs from different vendors. This allowed the implementation of a P25-based LMR network that consisted of RFSSs from multiple vendors. Although ISSI was originally meant for integrating two different P25 systems, the PTT application vendors have leveraged it as a way of moving radio traffic between a P25 system and a PTT service.

Integration of OTT PTT into LMR via ISSI is available. A crucial problem with OTT PTT is its lack of interoperability. For example, a user of Harris BeOn PTT cannot directly communicate with a Motorola Wave PTT's user; everyone needs to have the same application.

PoC services from Sprint, Verizon, and AT&T provide integration into LMR and feature rich PTT services. Verizon's PoC service, PTT+, offers ISSI-based LMR interworking via a virtual private network connection. These PoC services, however, do not provide interoperability between solutions from different vendors or even within the same vendor if the PoC service is provided by different carriers. This means that a user of AT&T's and a user of Verizon's PoC services cannot communicate with each other over PoC. First responders will benefit from using a globally recognized MCPTT implementation from 3GPP when it becomes available in future rather than an OTT PTT application or a PoC service.

If an LTE network running OMA-PoC as the native PTT protocol needs to interoperate with an existing P25 network, a P25 to OMA-PoC interoperability gateway is required that employs ISSI as the interworking protocol between the P25 network and the gateway. AT&T's Enhanced PTT or EPTT service is based on OMA's PCPS and is powered by Kodiak's PoC technology. It supports interoperability with LMR and allows AT&T EPTT subscribers to communicate with LMR radios by using an ISSI-based interoperability gateway. The ISSI gateway provides fast and reliable setup of communication sessions between the EPTT solution and LMR system.

8.3. Radio over IP (RoIP). *Radio over IP* (RoIP) is an application of Voice over IP (VoIP) technology to a two-way radio network. It is a generic term and does not describe any specific implementation or standard. RoIP is an expansion of the use of VoIP with additional control functions needed in LMR systems such as PTT.

RoIP offers a low-cost and reliable solution that meets the basic needs for PoC-to-LMR interoperability. RoIP has limited capabilities; for example, it does not support the passing of device IDs between networks. RoIP relies on a donor radio for connecting LMR and LTE networks. The advantage of the donor radio approach is that any radio technology (P25, TETRA, etc.) can be supported. However, RoIP systems require one donor radio for each LMR radio channel to be shared.

PoC vendors have integrated with RoIP systems to bridge their PoC systems to the LMR world. RoIP gateways can be connected to PoC servers that are hosted locally or in the cloud. Two types of RoIP gateways are available for providing interoperable PTT voice services between LTE and LMR including vendor-specific and third-party vendor-agnostic solutions. Motorola's WAVE and Harris' BeOn are examples of vendor-specific RoIP, while vendor-agnostic RoIP solutions include Mutualink and ESChat. JPS Interoperability Solutions also provides a modular vendor-agnostic RoIP gateway with multiple RoIP interfaces for PoC to LMR interoperability [31]. Figure 8 illustrates AT&T's EPTT system connected to an LMR system via JPS's RoIP gateway. Note that an RoIP gateway can connect any LMR talkgroup to a PoC talkgroup whereas the entire LMR network can be connected to the PoC network through an ISSI gateway.

9. Current Research Issues

3GPP has added additional mission-critical services in Release 14 that was completed in 2017. They include *Mission-Critical Data* (MCData) [32] and *Mission-Critical Video* (MCVideo) [33] over LTE. The MCData and MCVideo specifications offer equipment vendors and network operators a detailed set of standards that are available for implementation.

In Release 15, which is the first release of the 5G system, 3GPP is presently evaluating further mission-critical related topics including

(i) interworking between 3GPP defined mission-critical system and legacy mission-critical systems like P25 for voice and Short Data Service (SDS)

(ii) video push

(iii) video pull

(iv) off-network File Distribution Service (FDS)

(v) data streaming.

3GPP is also evaluating the ability for Unmanned Aerial Vehicles (UAVs) to be served as UEs using LTE network deployments in Release 15, which could be vital to employing UAVs for public safety operations in an LTE-based public safety network. The objective of the study in TR 36.777 [34] is to investigate the capability of LTE to provide connectivity to UAVs as aerial UEs.

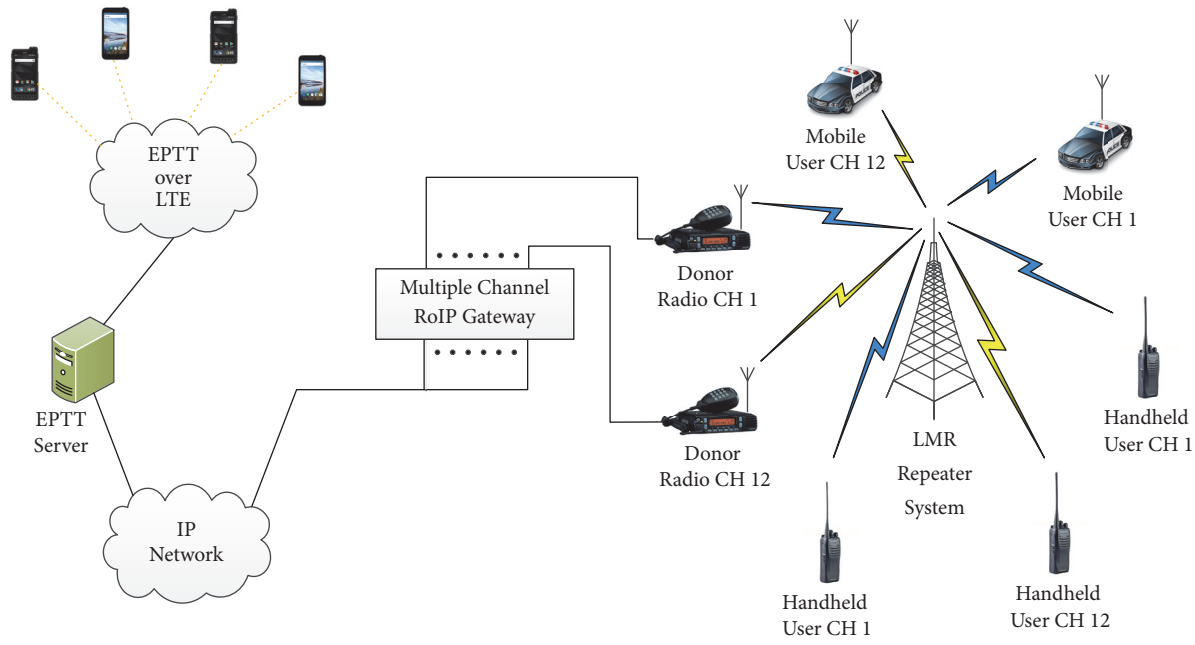


Figure 8: PoC to LMR interoperability via RoIP gateway [31].

The public safety oriented 3GPP specifications are titled MCPTT over LTE, MCData over LTE, MCVideo over LTE, etc. 3GPP is considering removing the "over LTE" text so that the same specifications can be reused for 5G work. Removing the LTE limitation from stage one requirements will help evolve the mission-critical services in stages two and three to support the 5G network.

9.1. Mission-Critical Data (MCData). The MCData service supports communication between a pair of users (i.e., one-to-one communication) and several users (i.e., group communication), where each user is able to

(i) share data using SDS

(ii) share files using FDS.

SDS is provided in both on-network and off-network modes while FDS is provided only in on-network mode.

SDS can deliver messages over the signalling channel or over a media bearer. SDS over media bearer can be standalone where the media bearer is set up for the purpose of delivering one message only and then disconnected afterwards or the media bearer can be established as a session for a group to carry multiple messages among group members. As a short message service, there is no pre-check for permission to transmit.

FDS can be session-based when all recipients are required to make mandatory download, or http upload-based where the file is temporarily stored in the controlling MCData server followed by distribution of notification of the file availability to the target recipients. In the latter case, the distribution control is managed and the recipient clients can manage their own reception. There is also provision for mandatory download with this method.

9.2. Mission-Critical Video (MCVideo). MCVideo defines a service for mission-critical video communication using LTE transport networks. Although this service is designed for transport over commercial and dedicated LTE networks, it is not expected to be limited to use over LTE.

MCVideo service includes

(i) video capture and encoding of the video information

(ii) secure streaming and storing of the video information

(iii) video decoding and rendering of the video information

(iv) processing of the video information, including the ability to annotate video frames and recognize video features

(v) mission-critical and public safety level functionality (e.g., group sessions, affiliations, end-to-end confidentiality, and emergency type communications) and performance (e.g., low latency)

(vi) transmission and control of the parameters relevant to those functions

(vii) secure operation such that video information can be reasonably unimpeachable when used in evidentiary procedures

(viii) definition and configuration of MCVideo groups and applications

(ix) configuration of the MCVideo users' profiles and of the MCVideo UEs

(x) interoperability with other services and systems.

While the streaming of video is part of the MCVideo service, the non-real-time or offline transfer of a video clip stored as a file containing video data is covered by the MCData service. An MCVideo UE is a device that provides video acquisition (e.g., has a camera), video rendering (has a display), or both and normally also has some encoding/decoding, communication and storage capabilities.

9.3. LMR–LTE Interworking/Interoperability. As discussed in Section 8, 3GPP's IWF will adapt LMR data and signaling to MCPTT data flows for connecting a P25 ISSI to the public safety LTE broadband network. To interconnect MCPTT over LTE to local P25 LMR networks, every single ISSI connection to the IWF would need to be separately and securely connected, which will make management and cost of this implementation unfeasible. The use of an ISSI hub has been proposed as an efficient way to achieve this interworking as it will be cost-effective and easier to secure a single point of entry into local P25 LMR networks [18].

9.4. Unmanned Aerial Vehicles (UAVs). 3GPP has mainly focused on the use of UAVs as aerial UEs and is working on specifying enhancements to improve terrestrial LTE networks for providing connectivity to UAV UEs. On the other hand, research is also being carried out to explore the role of UAVs as aerial base stations that can be deployed on-demand to boost network coverage. The use of UAVs as aerial UEs or aerial base stations can be vital for public safety applications during natural or man-made disasters.

9.4.1. UAV UEs. In disaster situations, UAVs could be deployed as aerial UEs of an LTE-based public safety network to relay images or videos of the disaster-hit area to augment the situational awareness of first responders on the scene or where the building or infrastructure may be temporarily inaccessible due to a hazard such as an earthquake or a chemical spill. The performance of the LTE network in the presence of UAV UEs is evaluated in 3GPP TR 36.777 in the following three scenarios:

(i) Urban macrocell with UAV UEs

(ii) Urban microcell with UAV UEs

(iii) Rural macrocell with UAV UEs.

The main issue with a UAV UE in the sky is the generation of uplink interference to multiple neighboring cells/eNodeBs. This uplink interference can degrade the performance of the existing UEs on the ground if not properly controlled or mitigated. Understanding the impact of uplink as well as downlink interference in the presence of UAV UEs is a key objective of the study in TR 36.777. Other aspects of LTE that are investigated include mobility performance and UAV identification. After the completion of this study, 3GPP has started a follow-up work item to specify enhancements to improve the efficiency and robustness of terrestrial LTE network for delivering efficient connectivity solutions for UAV UEs [35].

9.4.2. UAV Base Stations. As mentioned earlier in Section 5, DACA systems can be employed to temporarily restore critical communications during disasters such as earthquakes, hurricanes, or tsunamis, when terrestrial communication facilities are damaged resulting in communication outages. The main requirement in such scenarios is to provide broadband communication between first responders and between first responders and victims. Due to their mobility, UAVs serving as aerial base stations are a good fit for providing emergency communications where needed through a quickly deployable low-cost communication infrastructure.

A scenario in which different types of UAVs, including balloons, quadcopters, and gliders, serve as Unmanned Aerial Base Stations (UABSs) to provide emergency broadband connectivity after a disaster is investigated in [36]. The potential benefits of UABSs in the post-disaster scenario are investigated by evaluating the improvements in capacity and coverage of an LTE network that are gained after the deployment of UABSs. Once deployed, these UABSs form new small cells to improve network coverage. Unlike the fixed LTE communication infrastructure such as macro- or microcells, the positions of the UABSs can be adjusted and optimized to achieve better network performance. It should be noted that 3GPP has not yet provided any specifications for the standardization of aerial base stations.

10. Enabling Technologies for Future Public Safety Networks

The virtualization technologies can have tremendous advantages for public safety wireless networks in terms of multitenancy, programmability, and flexibility. The virtualization technologies such as *Software-Defined Networking* (SDN) and *Network Functions Virtualization* (NFV) act as key enablers for a Software-Defined RAN (SD-RAN) that provides the platform for RAN slicing thereby allowing public safety operators to effectively share existing RANs deployed by commercial operators.

10.1. SDN and NFV. The principal aim of SDN is to separate the control and data planes to achieve a programmable network. A centralized SDN controller executes all control tasks to facilitate network configuration and management. NFV is a strategy for the virtualization of network functions. The conventional network functions implemented as standalone boxes of specialized hardware and software are transformed into software components or virtual network functions running on virtual machines in the cloud. By employing SDN and NFV in the RAN, these virtualization technologies serve as key enablers of SD-RAN, which in turn is the key enabler of *RAN slicing*. RAN elements can be used as *service* offered to multiple core networks; the physical resources (i.e., base stations) can be shared by abstracting and slicing them into virtual RAN resources. This can enable the RAN to be shared by multiple operators resulting in significant savings on infrastructure costs.

10.2. Software-Defined RAN. In active sharing of network resources discussed earlier in Section 5, mobile network operators can share RAN to provide services using part of the available resources. However, the requirements of operators in terms of radio resources can dynamically change based on the needs of their subscribers, making RAN management a significant challenge. Using a *Software-Defined RAN* that is based on virtual network functions, operators can introduce programmability in the RAN to greatly simplify the control and management of RAN operations.

An open-source SD-RAN platform, referred to as FlexRAN, is proposed in [37] that incorporates an application programming interface for separation of control and data planes in the RAN. It offers a flexible control plane that supports other RAN management applications to be built over the FlexRAN controller. It provides flexibility to dynamically realize various degrees of coordination among base stations. It is transparent to the UEs, enabling easier deployment and evolution.

10.3. RAN Slicing in LTE. In RAN slicing based on SDN, NFV, and cloud computing, a slice orchestrator has been proposed that is responsible for instantiating slices from the EPC down to the RAN. Based on the slice provider requests, it instantiates a slice by selecting appropriate virtual network functions, e.g., core network functions such as forwarding of packets, session management, mobility management, and security, that have been virtualized. Each slice can have its own SDN controller to set up communication paths and manage the traffic within that slice. Resources such as infrastructure and radio spectrum can belong to the same or different network operators [38].

In an architecture for RAN slicing in LTE proposed in [38], the slice orchestrator instantiates instances of the EPC by selecting a set of virtual network functions or physical network functions for an instance. These EPC instances are connected to the shared RAN/eNodeB via the classical S1 interfaces or the new interfaces proposed by 3GPP in TR 23.799 [39]. By using the slice ID indicated by a UE that is hard encoded in the UE's SIM, the RAN/eNodeB steers the slice traffic to the appropriate EPC instance. Also, the eNodeB uses the slice ID to provide the appropriate slice resources such as the appropriate MAC schedular instance to satisfy the Quality of Service (QoS) required by a UE.

In the structure of the eNodeB under this solution, an abstraction layer in the form of a Resource Mapper (RM) is added. It provides an interface between the shared physical resource blocks and the Slice Resource Manager (SRM). Each slice has its own SRM that is responsible for scheduling UEs in its slice over virtual resources or virtual resource blocks. The RM maps the virtual resource blocks to physical resource blocks according to the amount of resources allowed to each slice. This RAN slicing architecture assumes that the slice orchestrator is employing an application programming interface such as FlexRAN to configure eNodeBs in real time.

10.4. RAN Slicing for Public Safety. To limit the time and cost of deployment as well as guaranty public safety requirements, a new communication architecture can be envisaged that introduces the notion of integrating a shared RAN with a dedicated one. Such a shared RAN is based on the concept of resource virtualization and RAN slicing. The architecture of a public safety network employing this concept is proposed in [40] and is illustrated in Figure 9, where RAN resources are shared with commercial networks. It is based on the concept of *multi-operator core network*—this is one of the network sharing configurations defined by 3GPP [41] in which only the RAN is shared.

The public safety and commercial networks maintain separate EPCs. *Public Safety EPC* (PS-EPC) has two connections via separate S1 interfaces with shared-RAN and dedicated-RAN whereas *Commercial EPC* (C-EPC) is linked with shared-RAN only as dedicated base stations are unable to support commercial UEs. The shared-RAN steers the traffic toward the appropriate EPC over the corresponding S1 interface based on a slice ID encoded in UE's SIM. The slice ID is also used over the X2 interface between the *Shared Base Station* (S-BS) and *Dedicated Base Station* (D-BS) for handover purposes.

The RAN management in the shared base station administers the resources in the shared-RAN. The Layer Resource Manager (LRM) coordinates among the shared base station and dedicated base stations in its coverage area for resource usage. Two slices are considered: commercial (*C-Slice*) and public safety (*PS-Slice*). Each S-BS comprises two slice resource managers to schedule respective UEs over virtual resources. A resource manager finally translates the virtual resources to physical resources. With this architecture, a public safety operator can exploit the nationwide RAN of a commercial operator based on appropriate service level agreements, while increasing access point density in few desired areas by deploying extra D-BSs.

11. Conclusions

First responders around the world are still largely using narrowband voice-centric communication networks such as LMR. These networks are designed to provide wide coverage and make extensive use of repeaters and device-to-device communications. The ease of use, reliability, and familiarity of these LMR networks will ensure that they will remain in existence for a long time to come. The primary aim of introducing a broadband public safety network is to take advantage of all recent radio technology advances. Also, a single nationwide wireless network for all public safety agencies can overcome critical issues like interoperability failures in current narrowband public safety networks.

LTE is a natural choice as a platform for introducing a range of new services to first responders. An LTE-based broadband nationwide public safety network can either share an existing nationwide commercial LTE network via upcoming technologies like RAN slicing or can be a completely separate wireless network dedicated for public safety agencies such as FirstNet. However, current LTE services for public safety like PoC, OTT PTT, PTV, and PTX do not meet mission-critical requirements, and 3GPP has addressed this via several recent enhancements such as proximity services,

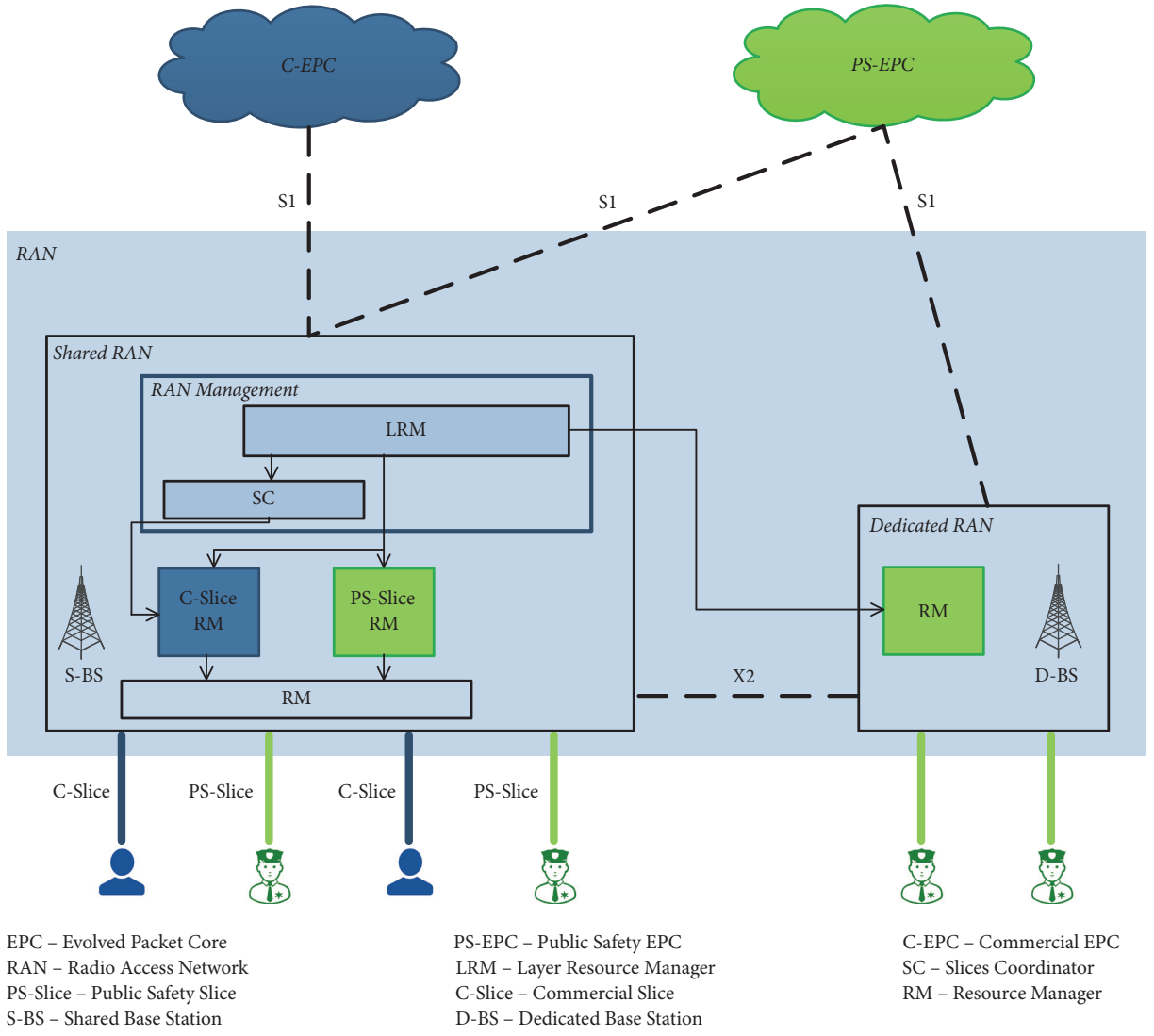


Figure 9: Network architecture employing RAN slicing [40].

group communication system enablers, and mission-critical PTT.

LTE will not replace LMR anytime soon and LMR is expected to continue to exist alongside any new LTE-based public safety network. However, LMR and LTE are quite different and getting them to interoperate effectively and seamlessly will need some innovative thinking. Existing interoperability solutions only provide integration of LMR with non-mission-critical LTE services like PoC and OTT PTT via ISSI and RoIP. Achieving standards-based interworking between legacy PTT on LMR and MCPTT on LTE will be critical. 3GPP has recently identified solutions suitable for interworking between LTE mission-critical systems and non-LTE mission-critical systems such as LMR.

Upcoming features and services of LTE such as mission-critical data, mission-critical video, and aerial UEs are also directly applicable to public safety. UAVs are also being considered as aerial base stations in LTE-based public safety networks that can be deployed on-demand to restore critical communications during disasters. In this paper, we provided a quick yet meaningful review of all these issues. We also offered a look ahead at the emerging virtualization technologies, such as SD-RAN, and RAN slicing, as enablers for future public safety networks.

A reasonable approach that is likely to dominate is to use the existing LMR voice-centric infrastructure as default mode of operation in an LMR-LTE integrated network. LTE is capable of providing a range of radio services that are not currently supported in LMR. For example, LTE can support radio bearers with variable QoS classes. Voice quality can be maintained even when the overall quality of the radio channel degrades. LTE can support rich multimedia environment through its IMS subsystem and MBMS. Also, LTE offers advanced network management capability with flexible access control. Other services will include execution of functions related to data analytics and mission-critical scene analysis. It is also likely that integrated LMR-LTE networks of the future will take advantage of upcoming

virtualization techniques such as SDN, NFV, SD-RAN, and RAN slicing. In the near future, LTE will be part of 5G and that will provide an even wider range of facilities including support for internet of things, virtual reality, and augmented reality. The interoperation of LMR and 4G/5G networks will also be imperative to the evolution and advancement of future public safety networks.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] A. Kumbhar, F. Koohifar, I. Güvenç, and B. Mueller, "A survey on legacy and emerging technologies for public safety communications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 97–124, 2017.

[2] T. Doumi, M. F. Dolan, S. Tatesh et al., "LTE for public safety networks," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 106–112, 2013.

[3] J. M. Peha, "Fundamental reform in public safety communications policy," *Federal Communications Law Journal*, vol. 59, no. 3, pp. 517–546, 2007.

[4] R. Ferrús, O. Sallent, G. Baldini, and L. Goratti, "LTE: the technology driver for future public safety communications," *IEEE Communications Magazine*, vol. 51, no. 10, pp. 154–161, 2013.

[5] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Toward moving public safety networks," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 14–20, 2016.

[6] A. Jarwan, A. Sabbah, M. Ibnkahla, and O. Issa, "LTE-based public safety networks: a survey," *IEEE Communications Surveys & Tutorials*, Article ID 2895658, 2019.

[7] A. Paulson and T. Schwengler, "A review of public safety communications, from LMR to voice over LTE (VoLTE)," in *Proceedings of the 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2013*, pp. 3513–3517, London, UK, September 2013.

[8] G. Baldini, S. Karanasios, D. Allen, and F. Vergari, "Survey of wireless communication technologies for public safety," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 619–641, 2014.

[9] Telecommunications Industry Association, "APCO Project 25 System and Standards Definition," Telecommunications Industry Association/Electronics Industry Association Telecommunications Systems Bulletins, TSB-102 Revision A, 1995.

[10] ETSI, "Terrestrial Trunked Radio (TETRA); Voice plus Data (V + D); Part 1: General Network Design," EN 300 392-1, V1.4.1, 2009.

[11] TETRAPOL Forum, "TETRAPOL Specifications; Part 2: Radio Air Interface," PAS 0001-2, V3.0.0, 1999.

[12] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2," Tech. Rep. TS 36.300, January 2016.

[13] 3GPP, "General Packet Radio Service (GPRS) Enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Access," Tech. Rep. TS 23.401, September 2014.

[14] 3GPP, "Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description," Tech. Rep. TS 23.246, June 2010.

[15] R. Hallahan and J. M. Peha, "The business case of a network that serves both public safety and commercial subscribers," *Telecommunications Policy*, vol. 35, no. 3, pp. 250–268, 2011.

[16] 3GPP, "Service Aspects and Requirements for Network Sharing (Release 11)," Tech. Rep. TR 22.951, September 2012.

[17] Federal Communications Commission, "The Role of Deployable Aerial Communications Architecture in Emergency Communications and Recommended Next Steps," White Paper, September 2011.

[18] Signals Analytics LLC, "Mission Critical Push-To-Talk (MCPTT) Implementation for Colorado," White Paper, December 2017.

[19] Motorola Solutions Inc., "PTT Technology Evolution," White Paper, June 2018.

[20] 3GPP, "Feasibility Study for Proximity Services (ProSe) (Release 12)," Tech. Rep. TR 22.803, August 2012.

[21] 3GPP, "Study on Architecture Enhancements to Support Group Communication System Enablers for LTE (GCSE_LTE) (Release 12)," Tech. Rep. TR 23.768, December 2013.

[22] 3GPP, "Proximity-based Services (ProSe); Stage 2 (Release 13)," Tech. Rep. TS 23.303, June 2015.

[23] 3GPP, "Group Communication System Enablers for LTE (GCSE_LTE) (Release 13)," Tech. Rep. TS 22.468, December 2014.

[24] 3GPP, "Group Communication System Enablers for LTE (GCSE_LTE); Stage 2 (Release 13)," Tech. Rep. TS 23.468, March 2015.

[25] 3GPP, "Mission Critical Push To Talk (MCPTT) over LTE; Stage 1 (Release 13)," Tech. Rep. TS 22.179, December 2015.

[26] Open Mobile Alliance, "Push to Communicate for Public Safety Control Plane," Tech. Rep. OMA-PCPS-TS-ControlPlane-V1_0-20150220-C, February 2015.

[27] Open Mobile Alliance, "OMA PoC Control Plane," Tech. Rep. OMA-TS-PoC-ControlPlane-V1_0-20060609-A, June 2006.

[28] National Public Safety Telecommunications Council LMR-LTE Integration and Interoperability Working Group, "Public Safety Land Mobile Radio (LMR) Interoperability with LTE Mission Critical Push to Talk," Tech. Rep., January 2018.

[29] 3GPP, "Study on Mission Critical Communication Interworking between LTE and Non-LTE Systems (Release 15)," Tech. Rep. TR 23.782, June 2017.

[30] Telecommunications Industry Association, "Project 25 Inter-RF Subsystem Interface Messages and Procedures for Voice and Mobility Management Services," Tech. Rep. TIA-102.BACA-A, January 2009.

[31] JPS Interoperability Solutions, AT&T Enhanced Push-To-Talk & Land Mobile Radio Integration via RoIP Technology from JPS Interoperability Solutions, https://jps.sugarmapleinteractive.com/wp-content/uploads/2016/04/EPTTLMR-Interoperability-via-JPS-RoIP.pdf, August 2016.

[32] 3GPP, "Mission Critical Data over LTE (Release 14)," Tech. Rep. TS 22.282, March 2017.

[33] 3GPP, "Mission Critical Video over LTE (Release 14)," Tech. Rep. TS 22.281, March 2017.

[34] 3GPP, "Study on Enhanced LTE Support for Aerial Vehicles (Release 15)," Tech. Rep. TR 36.777, December 2017.

[35] 3GPP, "New WID on Enhanced LTE Support for Aerial Vehicles," Tech. Rep. RP-172826, December 2017.

[36] A. Merwaday, A. Tuncer, A. Kumbhar, and I. Guvenc, "Improved throughput coverage in natural disasters: Unmanned aerial base stations for public-safety communications," *IEEE Vehicular Technology Magazine*, vol. 11, no. 4, pp. 53–60, 2016.

[37] X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. Kontovasilis, "FlexRAN: A flexible and programmable platform for software-defined radio access networks," in *Proceedings of the 12th ACM Conference on Emerging Networking Experiments and Technologies, ACM CoNEXT 2016*, pp. 427–441, Irvine, Calif, USA, December 2016.

[38] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on RAN: flexibility and resources abstraction," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, 2017.

[39] 3GPP, "Study on Architecture for Next Generation System (Release 14)," Tech. Rep. TR 23.799, December 2016.

[40] D. Marabissi and R. Fantacci, "Heterogeneous public safety network architecture based on RAN slicing," *IEEE Access*, vol. 5, pp. 24668–24677, 2017.

[41] 3GPP, "Network Sharing; Architecture and Functional Description (Release 10)," Tech. Rep. TS 23.251, June 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 8587932, 17 pages
https://doi.org/10.1155/2019/8587932

Research Article

Fast Multiattribute Network Selection Technique for Vertical Handover in Heterogeneous Emergency Communication Systems

Igor Bisio and Andrea Sciarrone

Department of Electrical, Electronic and Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Italy

Correspondence should be addressed to Igor Bisio; igor.bisio@unige.it

Received 25 January 2019; Accepted 25 March 2019; Published 16 April 2019

Academic Editor: Song Guo

The telecommunication infrastructure in emergency scenarios is necessarily composed of heterogeneous radio/mobile portions. Mobile Nodes (MNs) equipped with multiple network interfaces can assure continuous communications when different Radio Access Networks (RANs) that employ different Radio Access Technologies (RATs) are available. In this context, the paper proposes the definition of a Decision Maker (DM), within the protocol stack of the MN, in charge of performing network selections and handover decisions. The DM has been designed to optimize one or more performance metrics and it is based on *Multiattribute Decision Making* (MADM) methods. Among several MADM techniques considered, taken from the literature, the work is then focused on the TOPSIS approach, which allows introducing some improvements aimed at reducing the computational burden needed to select the RAT to be employed. The enhanced method is called Dynamic-TOPSIS (D-TOPSIS). Finally, the numerical results, obtained through a large simulative campaign and aimed at comparing the performance and the running time of the D-TOPSIS, the TOPSIS, and the algorithms found in the literature, are reported and discussed.

1. Introduction

Natural disaster events such as earthquakes, hurricanes, and floods and man-made ones such as terrorist attacks and toxic waste spills are facts of these last few years. Not only is the number of the disaster occurrences increasing very seriously, but also the number of killed and severely injured people is dramatically high. Environmental impacts on human health and quality of life are very high and there is an increase in disaster risk. Disaster management cycle is very complex and goes from prevention and mitigation to alert, response, and recovery. A quick deployment of a telecommunication infrastructure is essential in this case. For this motivation the role of the Information and Communication Technology (ICT) is topical for the deployment of a telecommunication infrastructure for risk and emergency management, for the technology integration, and for the implementation of specific technological developments possibly offering quick reconfigurability, interoperability, and scalability. The reference telecommunication infrastructure for the aforementioned scenario is necessarily composed of heterogeneous radio/mobile portions. On the other hand, nowadays the increased diffusion of mobile terminals, often called Mobile Nodes (MNs), has multiple network interfaces, such as Wi-Fi, WiMAX, and UMTS, and can assure consistent improvements in the mobile communication field. As a matter of fact, such a type of MN can operate inside the previously mentioned heterogeneous scenario where different Radio Access Networks (RANs) that employ different Radio Access Technologies (RATs) are available. Consequently, users can access a new set of services independently of their positions. It is worth noticing that different RATs have different characteristics and, of course, different strong points and weakness. These can also depend on the emergency conditions in which the MN is operating.

In the introduced context, a useful challenge is to assure an ubiquitous connection to MNs, defining an *anytime* and *anywhere* network, using the most appropriate RAN available

that best fits users' requirements. Following the *Always Best Connected* (ABC) criterion (see [1] and references therein) an MN is constantly connected to the core network using a specific RAN that matches some predefined requirements. Moreover, this paradigm enables the MNs to be aware of the heterogeneous and dynamic context where it is, represented by the available RANs status, and to adapt its behaviour consequently during the communication.

In this framework two processes cover a crucial role: *(i)* the *handover* that is the redirection of the active connections from a RAN to another one and *(ii)* the *network selection* that is a decisional process that selects the RAN that the MN has to use, among the available ones. According to the IEEE 802.21 standard, defined in [2] and described in Section 3, whose goal is to facilitate the interoperability of different Radio Access Technologies, these two functions are tightly linked: network selection is a fundamental part of the whole handover process as is detailed in Section 2.1. Both these functions are necessary to enable the MN to use the best RAN in terms of communications performance but they have stringent time requirements in their execution as stated in [3]. So an algorithm for the network selection must reduce the operations number necessary to perform the selection and, as a consequence, the needed time. On the contrary, an increase in the employed algorithm computational complexity may have a negative effect on the considered handover process that must wait for the selection procedure. This waste of time can provoke a performance detriment of the QoS perceived by the user and, in the worst case, can determine a service interruption.

During the network selection process, each RAN can be evaluated considering several characteristics often called attributes. Considering the mobile scenario, it is possible to say that the values of some attributes change dynamically while other attributes keep their values constant independently of the MN position inside the considered coverage area. As a consequence, as presented in [4], the attributes may be divided into 3 groups: *static*, *dynamic*, and *semidynamic*. We propose a novel formulation of the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) algorithm, taken from the Multiattribute Decision Making (MADM) theory [5], called Dynamic-TOPSIS (D-TOPSIS), which has the goal of limiting the operations number necessary to select the network, generating the same selection of the TOPSIS standard formulation.

The main contributions of this paper are listed below:

(i) A description of the mobile communications framework, regarding in particular handover process, network selection, and a possible applicative scenario, is included in Section 2.1.

(ii) A brief overview of the Standard IEEE 802.21, which plays a fundamental role in the management of heterogeneous RANs, is presented in Section 3.

(iii) A classification of the network selection algorithms, found in the literature, is reported in Section 4.

(iv) The formulation of both TOPSIS and D-TOPSIS algorithms, aimed at highlighting the difference among them and the reduction in the computational complexity, constitutes Section 5.

(v) The definition of a *Decision Maker*, within the protocol stack of the MN and in charge of performing network selection and handover decision, is proposed in Section 5.3. In this section the reference scenario used as a common test-bed to compare the considered algorithms is also described.

(vi) The numerical results, obtained through a large simulative campaign and aimed at comparing the different performance and the different execution time of the D-TOPSIS, the TOPSIS, and the algorithm found in the literature, are discussed in Section 6.

2. The Vertical Handover

The handover process can be divided into two categories: (i) *horizontal handover* and (ii) *vertical handover*. While the former occurs when the user switches between the same technologies RAT (for example, passing from a Wi-Fi network to another), the latter happens when the user moves from a technology to another. A possible example of vertical handover is the transparent switch from WiMAX to LTE.

Three main contributions are included in this section: a possible applicative scenario, the description of the handover process, and finally the presentation of the network selection concept.

2.1. A Heterogeneous Communication Network Example. In Figure 1 a possible heterogeneous network composition is represented. Suppose that a member of a rescue team operating during an emergency event is connected to a remote host, for example, an Emergency Operation headquarter, through an MN equipped with several heterogeneous network interfaces. The MN is moving by following the trajectory represented by black, dotted line with the arrow. Until he is in his house, the MN uses the domestic Wi-Fi connection. When he leaves the house, moving outside for the rescue operations, the Wi-Fi coverage area, the MN perceived that the quality of the Wi-Fi channel is degrading. Consequently, the MN decides to execute a handover and redirect the traffic to another RAN to prevent service interruption. In particular a UMTS and a WiMAX networks are available: in this case the network selection algorithm, implemented in the MN, selects the WiMAX network. Obviously, all these functions are executed automatically by the MN that maintains active the connection with the remote host, and the user is unaware of the handover execution. While the user is walking along his path, the MN periodically controls the state of the system, represented by the value of the attributes used to evaluate each RAN, and executes the network selection algorithm. Until the selected network is in use (represented by the black circles in Figure 1), the MN does not perform any handover. On the contrary, when the selected network has not been employed, the MN does the handover, for example, switching from WiMAX to UMTS as reported in Figure 1. Finally, when the user arrives at his office, the MN detects a Wi-Fi network that, according to the network selection algorithm, is

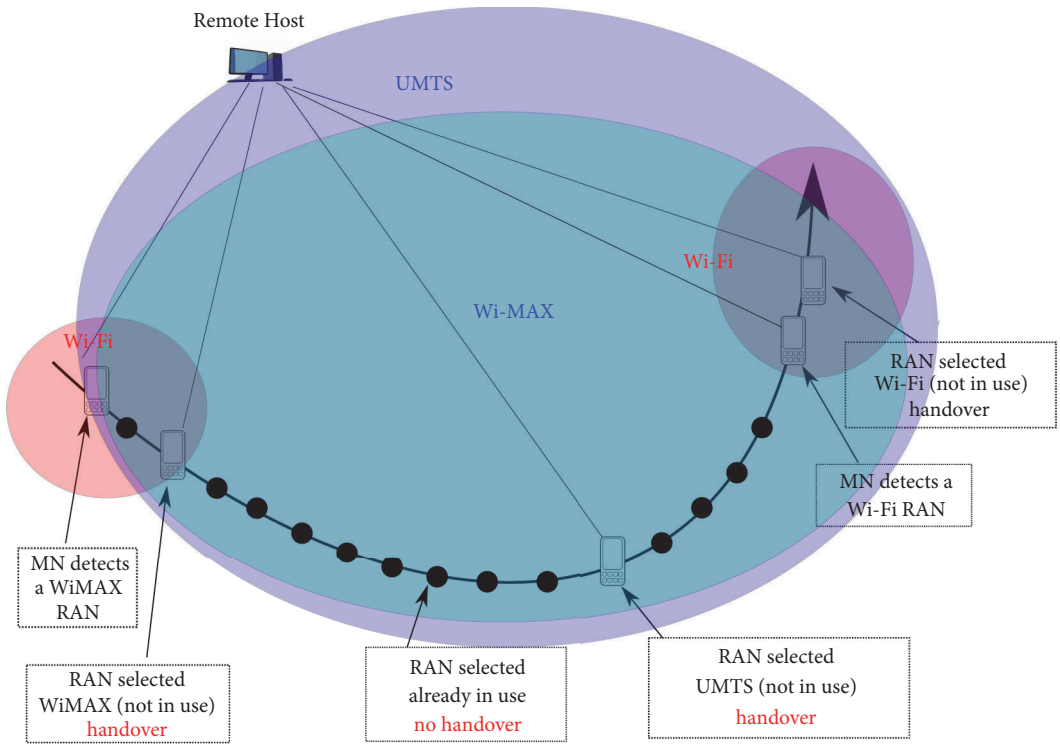


Figure 1: A possible applicative scenario of mobile communications.

better than the WiMAX in use, and it executes the handover. Similar behaviour characterizes also different scenarios such as the Intelligent Transportation Systems applications, which constantly monitor a vehicle moving along its path or inside a port area or in a construction site. Obviously, each network selection algorithm can give different decisions and it is characterized by different performance depending on the speed of the MN, as reported in Section 6.

The employment in this reference network of Wi-Fi, WiMAX, and UMTS technologies represents an example. The use of more modern and efficient mobile networks, such as LTE/4G and 5G, is object of ongoing performance investigation but it is worth noting that the network selection techniques presented and evaluated in this paper work independently of the available RATs.

2.2. The Handover Process. The term handover identifies the overall process that enables a Mobile Node (MN) to change the access network that it is using and to switch to another one among the available RANs. As report in [6] and briefly reprised at the beginning of this section, two different types of handover are defined according to the network technologies: *(i) horizontal handover*, when the RAT selected is the same one currently used by the MN, or *(ii) vertical handover*, when the target RAT is different from the one used by the MN. A typical example of horizontal handover happens when an MN moves outside a cell of a RAT such as the UMTS or the Wi-Fi. To maintain the communication active, the MN changes the Point of Access (PoA) and connects itself to the radio base station of the cell in which it enters. On the other hand, a vertical handover occurs, for example, when an MN exits the coverage area of a Wi-Fi network and switches to the UMTS technology in order to maintain the connection active.

Another classification of the handover process is proposed: *(i) soft handover* or *make-before-break* that represents the case in which an MN is connected to both PoAs during the switching phase and *(ii) hard handover* or *break-before-make* in which an MN can have contact with only one PoA at a time. A further type of handover, called *seamless*, meets the following requirements: *(i)* it is a soft handover, which means that the communication is maintained active during the network switching; *(ii)* the whole handover process has a limited duration; *(iii)* the handover must not determine any Quality of Service (QoS) degradation. Such a type of handover execution is also called transparent because the user is unaware of the RAN change performed by the mobile terminal.

2.3. The Network Selection Problem. The network selection is a task that is in charge of selecting the most appropriate RAN that an MN must use among the available ones. This procedure is tightly linked with the handover process. In fact, the decision taken by the network selection algorithm may trigger the handover execution when the selected network is not already used by the MN. In other words, the network selection is in charge of determining when the MN has to switch from a RAN to another: an MN may execute a handover not only when it moves away from the cell in use but also when a different available RAN is better from the network selection algorithm view point. In order to support the cooperation among different RATs the network is in charge of maintaining an efficient signalling architecture,

efficiently managing the resource utilization, and assuring the security and the integration of the communications. On the other, hand the network selection decision can be taken by the core network (i.e., *Centralized Network Selection*) or by the MN (i.e., *Distributed Network Selection*).

In a homogeneous environment, where there is only one type of RAT, the network selection is typically based on physical layer parameters, such as the Received Signal Strength Indicator (RSSI) or the Signal to Noise Ratio (SNR). In practice a Decision Maker (DM) entity evaluates only the quality of the channel and selects the PoA that assures the best channel conditions. These algorithms, which belong to the *single-attribute* group, are quite simple but they may provide a suboptimal selection from the user's viewpoint. In fact, different users may have different preferences and requirements not only referring to the quality of the communication but also, for example, to the power consumption minimization or to the Monetary Cost reduction. Moreover the various RATs, which characterize a heterogeneous scenario, may differ from each other not only in the channel quality but also in the aforementioned metrics. Consequently, another family of algorithms applied to solve the network selection problem is called *multiattribute* group; they are able to take advantages from the availability of different RATs and can select the best one evaluating more than one metric (or attribute) simultaneously [7]. More details about these groups of algorithms are reported in Section 4.

3. A Quick Overview of the IEEE 802.21 Standard

The IEEE 802.21 standard for the Media Independent Handover (MIH) defines extensible IEEE 802 media access independent mechanisms that enable the optimization of handover between heterogeneous IEEE 802 networks and facilitates handover between IEEE 802 networks and cellular networks, as reported in [2]. Its purpose is to assure to an MN the service continuity during the handover execution as well as after the handover. Moreover the standard is also aimed at assuring that the change of access network is not noticeable to the end user. As a logical consequence, it is mandatory not to decide the QoS degradation during the handover procedure limiting both packet losses and delay. It is worth noticing that two different aspects of QoS are considered by the standard: *(i)* the QoS experienced by an application during a handover and *(ii)* the QoS considered as part of a handover decision. Therefore, it is clear that this standard supports the handover execution not only for assuring the service continuity but also for selecting the RAN that best fits the MN requirements, in terms of QoS.

In more detail, it represents a common interface between the upper layers and the lower layers inside the protocol stack: even if an MN can have multiple network interfaces inside the lower layers, the MIH Layer is unique (for this reason, it is called technological dependent layer). A further logical element, defined by the standard, is a set of functions, called Media Independent Handover Function (MIHF), whose aim is to support the handover process. Indeed, this set of functions is in charge of activating the communication between the MIH Layer and the upper and the lower layers and providing the necessary information to support the handover procedures.

The standard defines three types of MIHF:

(i) *Media Independent Event Service* (MIES) comprehends the functions that provide the higher layers with some information sent by the technological dependent layer.

(ii) *Media Independent Command Service* (MICS) includes all the functions that send instructions from the higher layer to the technological dependent layers.

(iii) *Media Independent Information Service* (MIIS) defines a set of functions that provides the mechanism for retrieving information and assisting the handover decision.

More details about the aforementioned MIHF can be found in [2, 8].

According to [2] the communication model proposed by the IEEE 802.21 standard is represented in Figure 2, directly taken from [2]. An important virtual entity is introduced: the Point of Service (PoS). The MIHF of each network node becomes a PoS for an MN if it communicates directly with the same layer inside the MN protocol stack. In other words, a PoS is a network element that is in charge of providing an MN with the necessary information to perform the handover. In Figure 2 it is possible to see that there are two network portions, the *client side* where the MN is located and the *network side* that includes the PoS and the non-PoS elements, which are the network nodes that communicate with the MN indirectly. It is possible to view that the PoSs are logically located inside the Points of Access (PoAs) currently in use, inside the target PoAs, and inside other network nodes that are not PoAs but can be, for example, a database that contains some information of the neighbour RANs. All these nodes communicate together directly or indirectly, inside a client-server communication model, where the MN is the client that requires some information to execute the handover, and the core network is the server that provides the necessary information. As previously said, the IEEE 802.21 proposes a new layer, the MIH Layer, inside the protocol stack of each node. The communication with the other layers is based on the concept of Service Access Point (SAP) as reported in Figure 3. Three different types of SAP are defined by the standard:

(i) *MIH-SAP*: This SAP enables the communication between the upper layers, called also MIH Users, and the MIH Layer. Only a single MIH-SAP is implemented in each protocol stack and it is also called Technological Independent SAP

(ii) *MIH-LINK-SAP*: This type of SAPs assures the communication that the MN implements between the MIH Layer and each Link Layer. This communication takes place through media-specific instances of MIH-LINK-SAP for each different Link Layer. As a consequence, this SAP is also called Technological Dependent SAP.

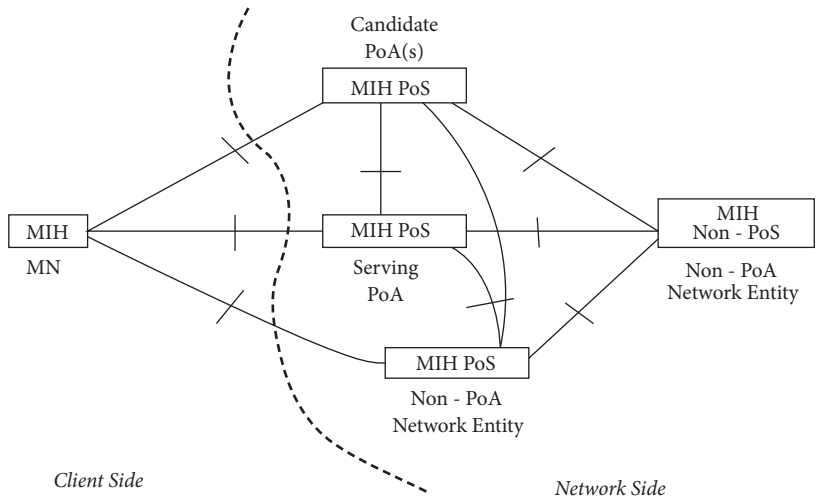


Figure 2: MIHF communication model proposed by [2].

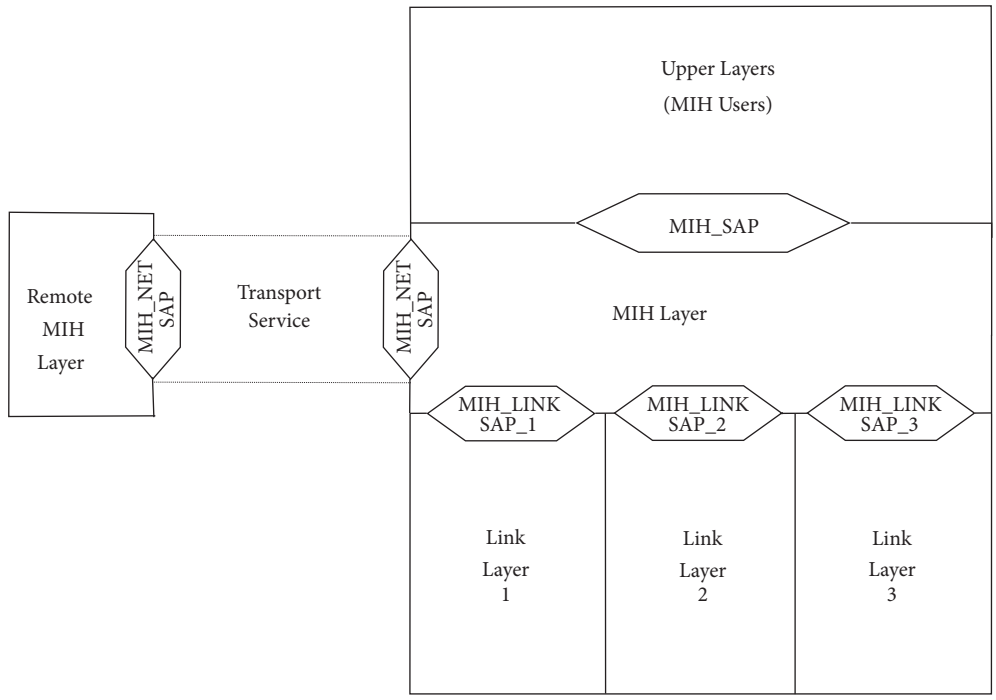


Figure 3: MIH Layer and SAPs reference model proposed by the IEEE 802.21 standard.

(iii) *MIH-NET-SAP*: This is the SAP that provides transport services supporting the exchange of MIH information and messages with the remote MIHF.

IEEE 802.21 also proposes the standardization of the handover procedure. The operations are divided into 3 phases, as reported in Figure 4:

(i) *Handover Initiation*: This is phase one of the handover process. The initiation comprehends message exchange with the Point of Access along with some preliminary measurements on the available RANs.

(ii) *Handover Preparation*: This is phase two. Here the Mobile Node chooses the network that will be employed after the handover and the negotiation for resource reservation that aims to grant that QoS requirements begin.

(iii) *Handover Execution*: In this final phase the traffic flows sent by the MN move to the selected RAN leaving the network access in use.

Some operations are included in the scope of the standard while others are only cited, but their implementation is not specified by the standard and, consequently, many different solutions can be applied. These operations are identified by an asterisk in Figure 4. Among these, a very important function is the handover decision, which refers to decisional process

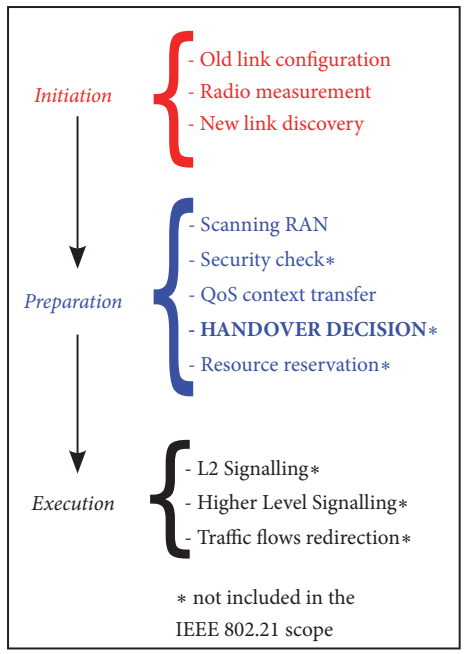


Figure 4: Three procedures that constitute the handover execution according to the IEEE 802.21 standard.

in charge of determining when the MN has to perform the handover and which is the target network: in practice, it identifies the network selection engine. The decision of the appropriate algorithm is something that has not been determined yet. Indeed, the algorithm that can be applied in this context has to assure good performance but at the same time it is characterized by stringent time constraints. It is worth noticing that a technique that selects the network may have a bad effect on the ongoing process, which remains in waiting status until the selection is performed.

4. The Network Selection Algorithms

The state-of-the-art comprehends two main branches of network selection algorithms: Single Performance Metric Optimization (SPMO) and MADM or, in the following, Multi-Performance Metrics Optimization (MPMO). SPMO algorithms choose the RAN by taking into account just one metric. MPMO algorithms take their decisions based on multiple metrics, at the same time, to choose the RAN to use. The object taking decisions about the RAN to be chosen is named Decision Maker (DM).

4.1. SPMO Algorithms. Each available network represents an alternative way that is validated trough one performance metric, named attribute. The DM is used to choose the best possibility, among the available ones, in accord with the maximization (or the minimization) of the metric, used as a utility (Cost) function. Let m be the possible RANs; we call the *alternatives* with an array $\mathbf{A} = (A_1, \ldots, A_j, \ldots, A_m)$. The j-th alternative is defined as $A_j = (x_j)$, where x_j is the considered metric (or attribute). The best option is denoted as A^{opt}_{SPMO} and is achieved by using the following equation.

$$A^{opt}_{SPMO} = \left\{ A_j : \arg\max_j \left(x_j \right) \right\} \tag{1}$$

Equation (1) is true if the employed metric needs to be maximized. If a metric needs to be minimized the $\arg\min_j(\cdot)$ operator is applied in (1).

Within the set of the possible metrics employed by the SPMO algorithms, a commonly taken decision is the RSSI, employed in [9, 10], which represents the received power measured in [dB] and is the parameter taken as a reference during the horizontal handover. Similarly, the same criterion could be employed for vertical handover. The method used for selection is very simple: the MN senses the RSSI from the PoAs among the RANs and determines the one with the highest value. Another weak point of this technique is the "Ping Pong effect" where there is repetitive and not useful handover, which occurs also if the RSSI value of an alternative RAN is slightly higher than the one related to the RAN currently employed [9]. This represents a drawback that affects negatively the QoS of MNs. The SPMO family techniques require a low computational burden, a low running time, and a limited power consumption. Differently, they can yield scarce performance if the goal would be the optimization of multiple metrics.

Target collision probability is another possible metric that has been used in network selection algorithms. Reference [11] uses it to maximize secondary users' throughput by employing a Markov queuing model. The authors evaluate their proposal by comparing it with other state-of-the-art strategies (Random and Greedy algorithms) showing its superior performance.

4.2. MPMO Algorithms. More than one metric is considered at the same time [7, 12]. They can be divided into

(1) QoS-based metrics, for example, packet loss, RSSI, bandwidth throughput, and transmission rate;

(2) power saving-based metrics, for example, MN battery lifetime and power consumption;

(3) other parameters-based metrics, for example, user preferences, Monetary Cost, and RAN security level.

DM task is to choose the best option in accord with a certain criterion. Most of the time MPMO techniques are characterized by higher computational burden than the SPMO ones. This is due to the fact that they can optimize simultaneously more metrics. MPMO approaches may select a suboptimal RAN by taking into account just one parameter and by finding the optimal solution considering all the different metrics at the same time.

Recalling the concept that the alternatives array m is defined with $\mathbf{A} = (A_1, \ldots, A_j, \ldots, A_m)$, for the MPMO algorithms the j-th alternative is denoted as $A_j = (x_{1j}, \ldots, x_{ij}, \ldots, x_{n,j})$, where x_{ij} represents the i-th attribute of the j-th alternative. The quantity n stands for the attributes number employed to compute each alternative.

4.2.1. Simple Additive Weight (SAW). The solutions that belong to this category (see [13] and references therein) leverage

on a quantity, called *cost*. In particular they assign to each *alternative* a *cost* value calculated trough the summation of the normalized value of each considered attribute. To further strengthen the importance of each attribute, weights may also be employed. Finally, the chosen network is the one that presents the smallest *cost* value. The mathematical detail is provided in (2).

$$V_{SAW}\left(A_j\right) = \sum_{i=1}^{n} w_i \cdot V_{SAW}^{A_j}\left(x_{ij}\right)$$
$$A_{MPMO-SAW}^{opt} = \left\{A_j : \arg\min_j \left(V\left(A_j\right)\right)\right\} \quad (2)$$
$$j = 1, \ldots, m$$

(i) $A_{MPMO-SAW}^{opt}$ is the provided alternative that the algorithm SAW has chosen;

(ii) $V_{SAW}(A_j)$ is the quantity linked to the alternative j, denoted by A_j (i.e., the cost);

(iii) $V_{SAW}^{A_j}(x_{ij})$ is the normalized cost of the j-th alternative computed by considering the i-th attribute x_{ij};

(iv) w_i is the weight associated with the i-th attribute.

This idea is already present in several papers within the same field, for example, in [13–17]. Within such a plethora of works, valuable realization of this approach can be found in [14]. It minimizes the generic cost of the employment of j-th network, based on the power consumption, Monetary Cost, and the bandwidth made available. The paper in [13] publishes a network selection strategy which defines the cost of the j-th network by using the weighted sum of the available bandwidth (normalized) and the Received Signal Strength Indicator (RSSI).

4.2.2. Weighted Product Method (WPM). It provides each alternative with a cost computed through the product of the attribute values [18]. It permits avoiding the normalization required by the SAW method. The detailed mathematical expression is written in (3).

$$V_{WPM}\left(A_j\right) = \prod_{i=1}^{n} V_{WPM}^{A_j}\left(x_{ij}\right)^{w_i}$$
$$A_{MPMO-WPM}^{opt} = \left\{A_j : \arg\min_j \left(V_{WPM}\left(A_j\right)\right)\right\} \quad (3)$$
$$j = 1, \ldots, m$$

(i) $A_{MPMO-WPM}^{opt}$ is the choice made by the WPM technique;

(ii) $V_{WPM}(A_j)$ is the quantity linked to the j-th alternative A_j;

(iii) $V_{WPM}^{A_j}(x_{ij})$ is the value of the i-th attribute of the j-th alternative x_{ij};

(iv) w_i is the weight associated with the i-th attribute.

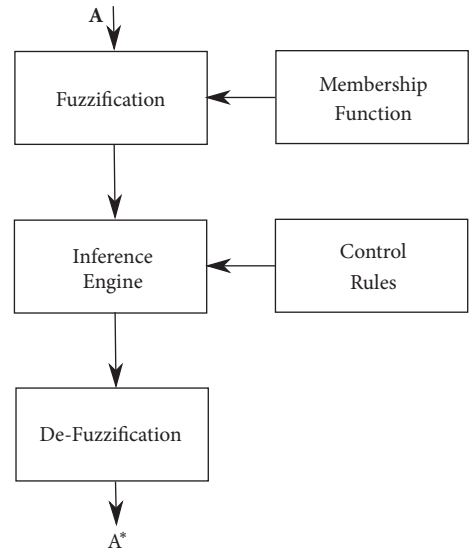


Figure 5: Block diagram of Fuzzy Logic Algorithm.

Apart from [18], WPM is not employed in many practical scenarios. For example, it is also used in [19] as a reference to show an analysis based on multiattribute error so as to obtain more robust differentiation between alternatives in a specific scenario.

Under certain constraints (see [20]), the minimization/maximization of the sum of logarithms is the same mathematical operation. Simplifying, $\max \prod_{j \in J} f_j(x)$ corresponds to $\max \sum_{j \in J} \ln f_j(x)$, where $f_j(x)$ are the utility functions. Reference [20] also stresses a similar concept.

To sum up, taking into account all the aforementioned reasons, the WPM has been taken as a reference in this paper.

4.2.3. Fuzzy Logic. A widely employed method for *network selection* consists in exploiting the Fuzzy Logic [21, 22]. It comes directly from the Fuzzy Set Theory. Specifically, the considered variables can have a "truth value" that can assume any value between 0 and 1. The scheme of a generic Fuzzy Logic technique is sketched in Figure 5. It takes as input all the alternatives **A**, evaluated according to the employed metrics. The algorithm begins with the *fuzzification* step that links RAN attributes to fuzzy sets by relying on each set *membership function*. Figure 6 provides an example of membership functions for three fuzzy sets. They are LOW, MEDIUM, and HIGH for the i-th attribute referred to as the j-th alternative, x_{ij}.

The next operation required by the Fuzzy Logic Algorithm is the use of the *inference engine*. Namely, in accord with the control rules, it decides a strategy to evaluate each RAN set of the attributes.

Finally, the ultimate step requires the *defuzzification*. The output provided by the *inference engine* will be employed to decide which is the best alternative.

4.2.4. Mixed Approach. The Mixed Approach merges together the Fuzzy Logic and a cost function for the *network selection*. This idea is reported in [23] in which 4 quantities

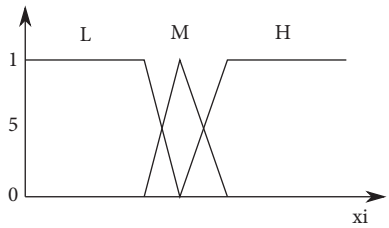


Figure 6: Membership function of the Fuzzy Logic Sets.

have been taken into account as input for the DM: Monetary Cost, available bandwidth, Received Signal Strength (*RSS*), and user preferences. Mixed Approach is then employed in order to decide the WWANs and the WLANs by making these steps:

(1) for each RAN, input quantities are normalization;

(2) for each parameter, the normalized values are to one fuzzy set: LOW, MEDIUM, HIGH. This is called *fuzzification*.

(3) Performance Evaluation Value (PEV) is finally computed and the best RAN is selected based on the highest PEV.

In the performance evaluation section, we refer to the Mixed Approach as Fuzzy-Simple Additive Weight (F-SAW). *Network Selection* methods that provide a similar approach can be found in [21, 24, 25].

5. The Proposed Network Selection Algorithm

The technique discussed in this paper, called D-TOPSIS, is a variation of the TOPSIS algorithm, already employed in the network selection algorithm [26–28]. This new formulation provides similar results in the decision processes and, at the same time, it lowers the operation number fundamental to obtain the solution.

5.1. The TOPSIS Algorithm. TOPSIS takes into account the *alternatives* defined by the quantities assumed by the considered *attributes*. The *i-th alternative* is denoted by an array $A_i = (x_{i1}^*, \ldots, x_{ij}^* \cdots, x_{in}^*)$ for $i \in [1, m]$ in which n and m are the number of *attributes* and the number of *alternatives*, respectively. As you can find in [5], the TOPSIS technique may be modeled by exploiting the geometry with m points in a n-dimensional space. Thus, we can use the Euclidean Norm to calculate the distance between each alternative and one or more reference points.

TOPSIS requires several steps that are summarized as follows:

(i) Calculation of the weighted normalized attribute values:

$$v_{ij} = w_j \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{4}$$

every $i = 1, \ldots, m$ *alternative*, for each $j = 1, \ldots, n$ *attribute*. w_j represents the weigh linked to the $j-th$ *attribute*, and the condition $\sum_{j=1}^{n} w_j = 1$ must hold.

(ii) Identification of the Positive Ideal Solution A^* (PIS) and the Negative Ideal Solution A^- (NIA), as detailed by

$$\begin{aligned} A^* &= \left(v_1^*, \ldots, v_j^*, \ldots, v_n^*\right) \\ &= \left(\left(\max_i v_{ij} \mid j \in J_1\right), \left(\min_i v_{ij} \mid j \in J_2\right) \mid i \right. \\ &\quad \left. = 1, \ldots, m\right) \\ A^- &= \left(v_1^-, \ldots, v_j^-, \ldots, v_n^-\right) \\ &= \left(\left(\min_i v_{ij} \mid j \in J_1\right), \left(\max_i v_{ij} \mid j \in J_2\right) \mid i \right. \\ &\quad \left. = 1, \ldots, m\right) \end{aligned} \tag{5}$$

in which J_1 is a quantity that stands for the set of positive attributes (which have to be maximized) and J_2 is a quantity the represents the set of negative attributes (which have to be minimized).

(iii) Calculation of the Separation Measures (SMs): to compute the distance between *alternatives* and the optimal, utopia point the Euclidean Norm is applied (see (6)).

$$\begin{aligned} SM_i^* &= \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2}; \quad for\ i = 1, \ldots, m \\ SM_i^- &= \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}; \quad for\ i = 1, \ldots, m \end{aligned} \tag{6}$$

(iv) Calculation of the Similarity Index (SI): for the $i-th$ *alternative*, A_i, the SI is calculated as $SI_i = SM_i^- / (SM_i^- + SM_i^*)$. The values range is within [0 – 1]. Specifically, the quantity $SI_i = 0$ occurs if the *alternative* coincides with the NIS (i.e., $A_i = A^-$). On the other hand, the case $SI_i = 1$ refers to the situation where the *alternative* coincides with the PIS (i.e., $A_i = A^*$). It is hence possible to state that the best *alternative* is represented by the one which presents the higher associated similarity index.

5.2. New Formulation of the TOPSIS Algorithm. The novel variant of the TOPSIS technique is named Dynamic-TOPSIS (D-TOPSIS). This name has the purpose to stress the concept that the decision at the generic step t also considers the decision previously taken. In more detail, at the step t employed to compute the $i-th$ *alternative*, the *attributes* are split into two groups: the *static attributes* $\mathbf{s}_i(t)$ and the *dynamic attributes* $\mathbf{d}_i(t)$. For each *alternative* the attributes keep their numbers constant at each step t when the network selection is performed, when the *alternative* is available. Consequently, for the $i-th$ *alternative* the quantity associated with each static *attribute* at the step t is equal to the value

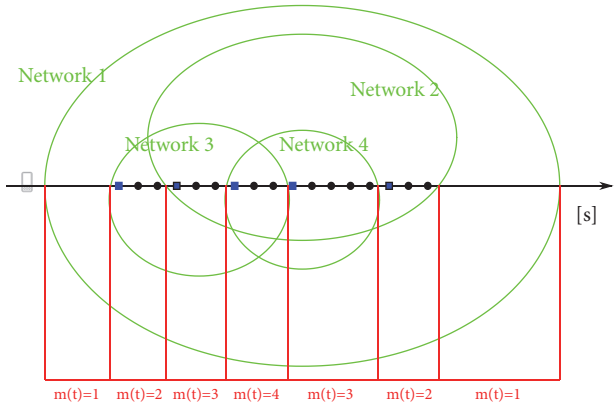


FIGURE 7: Variation of the number of alternatives over time in the network selection problem.

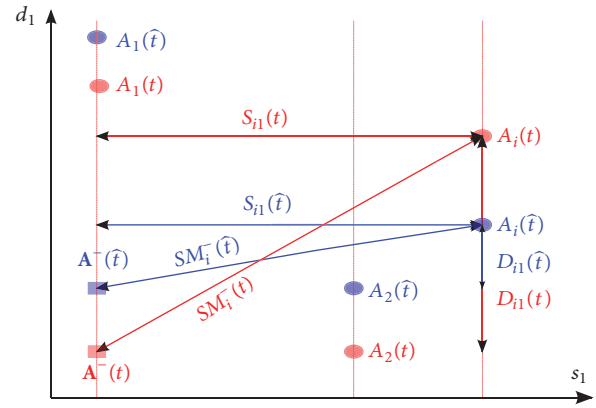


FIGURE 8: Distance of the $i-th$ alternative from the Negative Ideal Solution (NIS).

of the same attribute at the time of the previous TOPSIS execution $\hat{t}$. In practice $s_i(t) = s_i(\hat{t})$.

The next steps, which must be done to compute the selection, depend on the alternatives that are available at the instant t: if the set of available RANs is the same in t and in $\hat{t}$ the nonstatic (i.e., the dynamic) form of the technique is used, while, if this condition does not occur, the static version of the technique is then used. Figure 7 highlights how this condition impacts on the network selection formulation: the position of the MN moving along its path is represented. It enters and exits the coverage area of different RANs and periodically executes the network selection. The number of available *alternatives* (i.e., the number of available networks), $m(t)$, in each instant in which the selection is performed, is also reported. In particular, the blue squares represent the execution of the standard TOPSIS algorithm, while the black circles identify the execution of the D-TOPSIS.

As a consequence of the distinction among static and dynamic *attributes* the $i-th$ *alternative* can be defined as in

$$\mathbf{A}_i(t) = \left(\mathbf{s}_i(\hat{t}), \mathbf{d}_i(t)\right) = \left(s_{i,1}(\hat{t}), \ldots, s_{i,j}(\hat{t}), \ldots, s_{i,n_s}(\hat{t}), d_{i,j}(t), \ldots, d_{i,j}(t), \ldots, d_{i,n_d}(t)\right) \quad (7)$$

where n_s and n_d are the number of static and dynamic *attributes*, respectively. The matrix $\mathbf{A}(t)$, reported in (8), describes all the available *alternatives* at the step t.

$$\mathbf{A}(t) = \begin{pmatrix} \mathbf{A}_1(t) \\ \vdots \\ \mathbf{A}_i(t) \\ \vdots \\ \mathbf{A}_{m(t)}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1(\hat{t}), \mathbf{d}_1(t) \\ \vdots \\ \mathbf{s}_i(\hat{t}), \mathbf{d}_i(t) \\ \vdots \\ \mathbf{s}_{m(t)}(\hat{t}), \mathbf{d}_{m(t)}(t) \end{pmatrix} \quad (8)$$

As previously said the TOPSIS algorithm is based on the concept of distance between alternatives represented by points inside a multidimensional space. Applying the Euclidean Norm it is possible to measure the distance between each alternative and two reference points (the PIS and the NIS previously defined). In order to better explain the D-TOPSIS approach, the distance reported above is highlighted in Figure 8. Both the *attributes* are positive which means that they must be maximized. One of them is static (s_1) and the other is dynamic (d_1). In Figure 8 only the distance from the NIS is represented. Even though not plotted, analogous deductions can be obtained for the PIS case.

Considering Figure 8 the NIS $A^-(t)$ at the instant t and the NIS $A^-(\hat{t})$ at the instant $\hat{t}$ are denoted by a red and blue squares, respectively. The distance between the PIS and the $i-th$ *alternative* at the instant t is determined by the two components $S_{i1}(t)$ and $D_{i1}(t)$. In this paper, we name it *Partial Distance*. It have to be computed for each *alternative* for each run of the TOPSIS technique. From Figure 8 we can see that $S_{i1}(t) = S_{i1}(\hat{t})$. Thus, there is no need to compute the value of this parameter since it reduces the operations number necessary to effectuate the RAN selection. In an intuitive way, such reduction is even more notable if the number of static parameters or the number of *alternatives* increases.

As a consequence, the algorithm proposed is based on the following steps:

(i) Calculation of the weighted normalized values of the static and dynamic *attributes*.

$$N\left(s_{ij}(t)\right) = \begin{cases} N\left(s_{ij}(\hat{t})\right); & \text{if RANs do not change} \\ w_j \dfrac{s_{ij}(t)}{\sqrt{\sum_{i=1}^{m(t)} s_{ij}(t)^2}}; & \text{otherwise} \end{cases}$$
$$i = 1, \ldots, m; \quad j = 1, \ldots, n_s; \quad (9)$$

$$N\left(d_{ij}(t)\right) = w_j \frac{d_{ij}(t)}{\sqrt{\sum_{i=1}^{m(t)} d_{ij}(t)^2}}$$
$$i = 1, \ldots, m; \quad j = 1, \ldots, n_d;$$

(ii) Identification of the Positive $A^*(t)$ and the Negative $A^-(t)$ Ideal Solution

$$A^*(t) = \begin{cases} \left(\mathbf{N}\left(\mathbf{s}^*\left(\hat{t}\right)\right), \mathbf{N}\left(\mathbf{d}^*(t)\right)\right); & \text{if RANs do not change} \\ \left(\mathbf{N}\left(\mathbf{s}^*(t)\right), \mathbf{N}\left(\mathbf{d}^*(t)\right)\right); & \text{otherwise;} \end{cases} \quad (10)$$

where

$$\begin{aligned} \mathbf{s}^*(t) &= \left(N\left(s_1^*(t)\right), \ldots, N\left(s_{n_s}^*(t)\right)\right) \\ &= \left(\left(\max_i N\left(s_{ij}(t)\right) \mid j \in J_{s_1}\right), \right. \\ &\quad \left.\left(\min_i N\left(s_{ij}(t)\right) \mid j \in J_{s_2}\right)\right) \\ \mathbf{d}^*(t) &= \left(N\left(d_1^*(t)\right), \ldots, N\left(d_{n_d}^*(t)\right)\right) \\ &= \left(\left(\max_i N\left(d_{ij}(t)\right) \mid j \in J_{d_1}\right), \right. \\ &\quad \left.\left(\min_i N\left(d_{ij}(t)\right) \mid j \in J_{d_2}\right)\right). \end{aligned} \quad (11)$$

$$A^-(t) = \begin{cases} \left(\mathbf{N}\left(\mathbf{s}^-\left(\hat{t}\right)\right), \mathbf{N}\left(\mathbf{d}^-(t)\right)\right); & \text{if RANs do not change} \\ \left(\mathbf{N}\left(\mathbf{s}^-(t)\right), \mathbf{N}\left(\mathbf{d}^-(t)\right)\right); & \text{otherwise} \end{cases} \quad (12)$$

where

$$\begin{aligned} \mathbf{N}\left(\mathbf{s}^-(t)\right) &= \left(N\left(s_1^-(t)\right), \ldots, N\left(s_{n_s}^-(t)\right)\right) \\ &= \left(\left(\min_i N\left(s_{ij}(t)\right) \mid j \in J_{s_1}\right), \right. \\ &\quad \left.\left(\max_i N\left(s_{ij}(t)\right) \mid j \in J_{s_2}\right)\right) \\ \mathbf{N}\left(\mathbf{d}^-(t)\right) &= \left(N\left(d_1^-(t)\right), \ldots, N\left(d_{n_d}^-(t)\right)\right) \\ &= \left(\left(\min_i N\left(d_{ij}(t)\right) \mid j \in J_{d_1}\right), \right. \\ &\quad \left.\left(\max_i N\left(d_{ij}(t)\right) \mid j \in J_{d_2}\right)\right). \end{aligned} \quad (13)$$

In (11) and (13) J_{s_1} and J_{d_1} represent the sets of positive static and dynamic *attributes*, respectively, that have to be maximized. Similarly, J_{s_2} and J_{d_2} represent the sets of negative static and dynamic *attributes* that need to be minimized, respectively.

(iii) Calculation of the *Partial Distances* for all the attributes between each *alternative* and the Ideal Solutions as in (14) and (15).

$$N\left(S_{ij}^*(t)\right) = \begin{cases} S_{ij}^-\left(\hat{t}\right); & \text{if RANs do not change} \\ \left|N\left(s_{ij}(t)\right) - N\left(s_j^*(t)\right)\right|; & \text{otherwise} \end{cases}$$
$$i = 1, \ldots, m; \; j = 1, \ldots, n_s;$$

$$D_{ij}^*(t) = \left|N\left(d_{ij}(t)\right) - N\left(d_j^*(t)\right)\right|$$
$$i = 1, \ldots, m; \; j = 1, \ldots, n_d; \quad (14)$$

$$S_{ij}^-(t) = \begin{cases} S_{ij}^-\left(\hat{t}\right); & \text{if RANs do not change} \\ \left|N\left(s_{ij}(t)\right) - N\left(s_j^-(t)\right)\right|; & \text{otherwise} \end{cases}$$
$$i = 1, \ldots, m; \; j = 1, \ldots, n_s; \quad (15)$$

$$D_{ij}^-(t) = \left|N\left(d_{ij}(t)\right) - N\left(d_j^-(t)\right)\right|$$
$$i = 1, \ldots, m; \; j = 1, \ldots, n_d;$$

(iv) Calculation of the Separation Measures (SMs) as reported in (16).

$$\begin{aligned} SM_i^*(t) &= \sqrt{\sum_{j=1}^{n_s}\left(S_{ij}^*(t)\right)^2 + \sum_{j=1}^{n_d}\left(D_{ij}^*(t)\right)^2} \\ SM_i^-(t) &= \sqrt{\sum_{j=1}^{n_s}\left(S_{ij}^-(t)\right)^2 + \sum_{j=1}^{n_d}\left(D_{ij}^-(t)\right)^2} \\ & \qquad i = 1, \cdots, m; \end{aligned} \quad (16)$$

(v) Calculation of the Similarity Index (SI) as $SI_i(t) = SM_i^-(t)/(SM_i^-(t) + SM_i^*(t))$. The highest Similarity Index identifies the best alternative for the D-TOPSIS algorithm too.

Employing the D-TOPSIS algorithm for each *alternative*, it is not mandatory to compute both values of the *Partial Distances* for the static *attributes*. Indeed, these values are memorized during the last run of the standard TOPSIS and therefore loaded during the execution of the D-TOPSIS. As a consequence, the sole *Partial Distances* referred to as the dynamic attributes have to be determined so as to compute the Separation Measures and the Similarity Index, during the execution of each network selection process independently of the set of available alternatives.

5.3. The Network Selection as IEEE 802.21 Component. In order to compare the performance of the D-TOPSIS with the other network selection algorithms described in Section 4, the authors propose a definition of a new component, called *Decision Maker* (DM), modeled as a virtual entity, which is in charge of performing the selection of the RAN implementing the considered algorithms. The structure of the DM is represented in Figure 9, where it is possible to see that it is integrated into the MIHF Layer of the MN. As a matter of fact this virtual entity is deputy to take the handover decision functions, which is included in but not defined by the IEEE 802.21 standard, as reported in Section 3. In other words, the DM senses the heterogeneous environment in which the MN is moving; it acquires the characteristics of each RAN and,

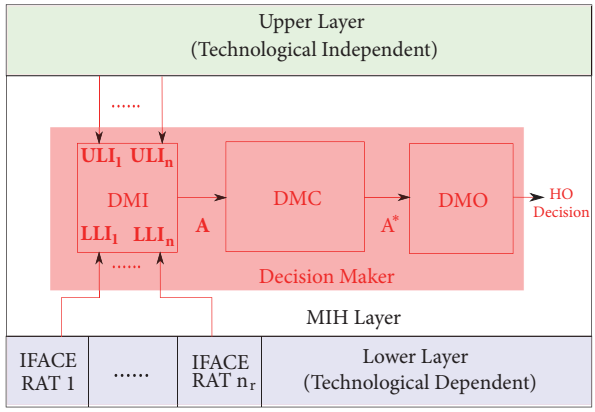


Figure 9: Structure of the proposed Decision Maker set inside the MIH Layer.

according to the implemented algorithm, selects the one that the MN has to use.

Observing Figure 9, it is possible to see that the DM is composed of three parts. The first block, called *Decision Maker Input* (DMI), collects the values of each metric that defines each available RAN through the MIH Layer primitives. This information belongs to two different groups: *(i)* Information from the technological dependent lower layers, which comprises the Lower Layers Information (**LLI**) vector, one for each available RAN. This information is taken directly from the heterogeneous RAT interfaces and includes, for example, the RSSI and the power consumption. It is worth noticing that n_r is the number of different RAT interfaces of the MN, in Figure 9. *(ii)* Information from the technological independent upper layers that comprises the Upper Layer Information (**ULI**) vector, again one for each available RAN. This information is collected inside the matrix **A** created by the DMI.

The algorithm used to evaluate the RANs available is implemented inside the *Decision Maker Core* (DMC) block. It receives as an input the matrix **A** used to describe each RAN (i.e., each alternative) and produces as an output the alternative selected according to the used algorithm, $\mathbf{A}^*$.

Finally the *Decision Maker Output* (DMO) block receives as input the selected RAN from the DMC block and it decides to perform the handover, if the selected network is not already in use; otherwise, it decides to do nothing. If it decides to change RAN, it initiates the handover procedure using the MIH Layer primitives.

6. Performance Analysis

6.1. Simulation Scenario. As already done in other works (see [29]), the simulated scenario adopted to test the network selection algorithms and to compare their performance has been realized by using Network Simulator 2 (*ns-2*). To support the IEEE 802.21 standard, the package provided by the National Institute of Standards and Technology (NIST) is used inside *ns-2* [30]. Moreover the authors implement in the simulated MN the aforementioned DM and the considered network selection algorithms.

Table 1: Range value of the attribute considered.

Parameters	Range Value
Power Consumption	[0, 16 - 0, 22] w
Monetary Cost	[1-10]
Wi-Fi Capacity	[1 - 20] Mbps
Wi-Max Capacity	2 Mbps
UMTS Capacity	0.384 Mbps

The communication is established by a User Datagram Protocol (UDP) traffic flow generated by a remote host to the MN. The DM inside the MN collects the values of each attribute of each RAN, executes the network selection algorithm, and decides to perform or not the handover each 5 [s] (i.e., the decision period) when two or more RANs are available.

As in the reference example described in Section 2.1, three different RATs are considered: UMTS, WiMAX (IEEE 802.16), and Wi-Fi (IEEE 802.11). There is a single UMTS cell that contains the entire environment equal to 2000x2000 [m], a single WiMAX cell, and 8 Wi-Fi cells. Two separate cases are taken into account according to the MN speed: *pedestrian*, 3 [m/s], and *vehicular*, 10 [m/s]. The overall duration of the simulation is set equal to 500 [s]. While the aforementioned parameters are kept constant in each simulation, other parameters are casually determined in each simulation. Such parameters are the position and the dimension of the WiMAX and Wi-Fi networks, the start point and the end point of the MN.

The metrics used as attributes of the RANs are four ($n = 4$); three of them are static ($n_s = 3$) and only one is dynamic ($n_d = 1$).

(i) The *Received Signal Strength Indicator* measure which is related to the distance between the MN and the PoA of each RAN. It represents a positive, thus having to be maximized by the network selection algorithm, and dynamic attribute.

(ii) The *Capacity* that each RAN reserves for the MN. It is a positive and static attribute. Its value, for each RAN, is set in each simulation according to Table 1.

(iii) The *Monetary Cost* that the MN has to pay to use a RAN. It is a negative, thus having to be minimized by the network selection algorithm, and static attribute. Its value, for each RAN, is set in each simulation, according to Table 1.

(iv) The *Power Consumption* of the MN to maintain the communication active with the remote host. It is a negative and static attribute. Again, its value, for each RAN, is set in each simulation, according to Table 1.

It is worth noticing that, in Table 1, the Monetary Cost is a rough indication that allows sorting the network from the cheaper, in which *MN* is set equal to 1, to the more expensive one, with $MC = 10$.

6.2. Performance Comparison. In the performance comparison, 9 network selection algorithms are considered.

Specifically, we have employed 5 MADM techniques that comprise also the TOPSIS and the D-TOPSIS. The rest are single-attribute Decision Maker algorithms. Namely, each of the remaining selection algorithms is devoted to the optimization of one of the attributes only: *(i)* Received Signal Strength Indicator based, *(ii)* Available Capacity, *(iii)* Monetary Cost, and *(iv)* Power Consumption.

To evaluate the network selection algorithms six metrics are adopted:

(i) Capacity (C) (expressed in [bps]) that the RAN in use assigns to the traffic flow transmitted by the remote host to the MN;

(ii) $RSSI$ (expressed in [dBW]);

(iii) Monetary Cost (MC) payed for employing the RANs;

(iv) power consumption (P) (expressed in [W]) of the MN to maintain the communication active;

(v) packets delay (D) calculated as the difference between the packet transmission time and the time in which the packet is received by the MN in [s];

(vi) number of handover processes executed by the MN (H).

All the aforementioned metrics are negative except for the first two. The optimal choice of the network selection technique is to select the network that guarantees the best compromise between the considered metrics, maximizing the positive one and minimizing the negative one.

In Figures 10, 11, 12, 13, 14, and 15 the evaluated techniques are labeled as follows: the single-attribute methods that optimize $RSSI$, C, MC, and P are reported as #1, #2, #3, and #4, respectively; the multiattribute optimization approaches SAW, *Fuzzy Logic*, WPM, TOPSIS, and D-TOPSIS are indicated with #5, #6, #7, #8, and #9, respectively.

Concerning the $RSSI$ metric, among the single-attribute methods, the best performance (i.e., the highest $RSSI$) is obtained, obviously, if the metric optimized is exactly $RSSI$ (in this case #1). The MADM techniques perform satisfactorily in all cases: from #5 to #9 the obtained $RSSI$ is almost equal to #1. In general the obtained $RSSI$ is approximately similar for all the evaluated network selection approaches. Analogous considerations are valid if the other performance metrics are used to evaluate the techniques.

Indeed, when MADM tools are applied, it is not simple to define a best performance: the outcome of a MADM solution is a compromise by definition. The key point is to obtain comparable performance with respect to the cases in which the metrics are considered singularly. This is true for almost all the considered MADM techniques employed in this paper to implement the network selection. The crucial difference is the time needed to find the aforementioned compromise solution: this motivates the analysis reported in the next subsection.

6.3. Computational Complexity Study. A computational complexity analysis of TOPSIS and D-TOPSIS is discussed in this subsection. As previously said, one of the most important requirements for an MN is to limit the complexity of the implemented algorithm in order to reduce the execution time. As a consequence, limiting the number of necessary operations to select a network is a crucial task in order to define a new algorithm.

Table 2: Number of operations performed by TOPSIS and D-TOPSIS.

Operation	Standard TOPSIS	Dynamic TOPSIS	Percentage
ArgMin	$n_s + n_d$	n_d	$\frac{n_s}{n_s + n_d}$
ArgMax	$n_s + n_d$	n_d	$\frac{n_s}{n_s + n_d}$
Multiplication	$m \cdot (n_s + n_d)$	$m \cdot n_d$	$\frac{n_s}{n_s + n_d}$
Power 2	$m \cdot (n_s + n_d) \cdot (m + 2)$	$m \cdot n_d \cdot (m + 2)$	$\frac{n_s}{n_s + n_d}$
Subtraction	$2 \cdot m \cdot (n_s + n_d)$	$2 \cdot m \cdot n_d$	$\frac{n_s}{n_s + n_d}$
Division	$m \cdot (n_s + n_d + 1)$	$m \cdot (n_d + 1)$	$\frac{n_s}{n_s + n_d + 1}$
Square root	$m \cdot (n_s + n_d + 2)$	$m \cdot (n_d + 2)$	$\frac{n_s}{n_s + n_d + 2}$

The first two columns of Table 2 show the number of operations that are necessary to select the *alternative* with the two TOPSIS algorithm variants. The third column indicates the reduction (as a percentage) in the number of operations obtained using the D-TOPSIS algorithm with respect to the TOPSIS. This percentage is computed as the difference between the number of operations for the D-TOPSIS and that for the standard TOPSIS, divided by the number of operations of the TOPSIS.

Each row identifies one operation used in the formulation of both the algorithm versions; as previously defined n_s is the number of static *attributes*, n_d is the number of dynamic *attributes*, and m is the number of available *alternatives* (i.e., the number of available RANs). It is important to notice that the percentage of reduction in the number of operations does not depend on the number of *alternatives* m that are available. This consideration is confirmed by the numerical results discussed in Section 6. Moreover the reduction in the number of operation is explicitly calculated considering different number of dynamic and static attributes (i.e., $n_d = [1, \ldots, 10]$ and $n_s = [1, \ldots, 10]$). In Figure 16 the reduction of a first group of operations is plotted, including the *Argmin*, the *Argmax*, *Multiplication*, *Power 2*, and *Subtraction*, which is equal to $n_s/(n_s+n_d)$ as reported in Table 2. Two further groups of operations are considered, referred to as the *Division* and the *Square root*. Their reductions, which are, respectively, equal to $n_s/(n_s+n_d+1)$ and $n_s/(n_s+n_d+2)$, are calculated over the same variation of the number of attributes considered for the first group and are plotted, respectively, in Figures 17 and 18. As previously said a network selection method has stringent execution time constraints in order to avoid unnecessary time waste for the handover process as a whole. As a consequence it is necessary to limit the computational complexity of the used technique. The proposed D-TOPSIS algorithm can assure great benefits in this direction with

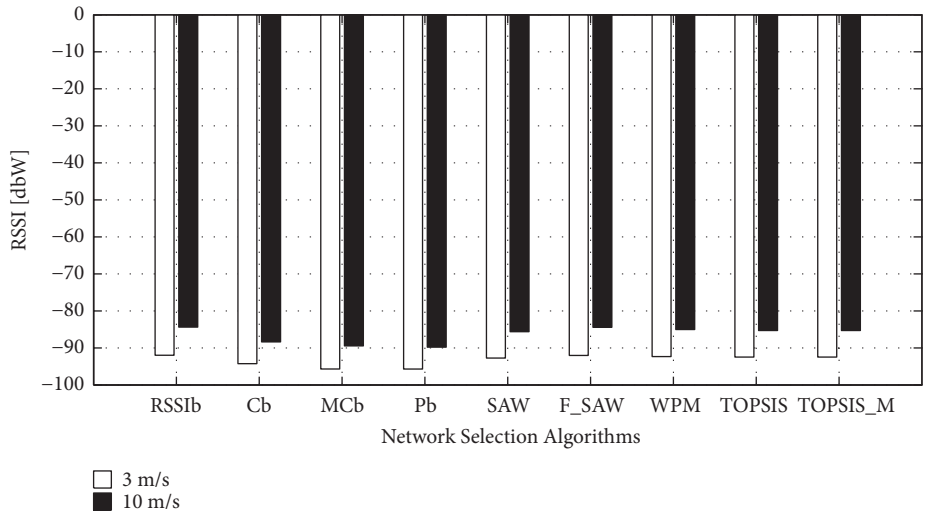


Figure 10: The considered *RSSI* metric for various network selection techniques.

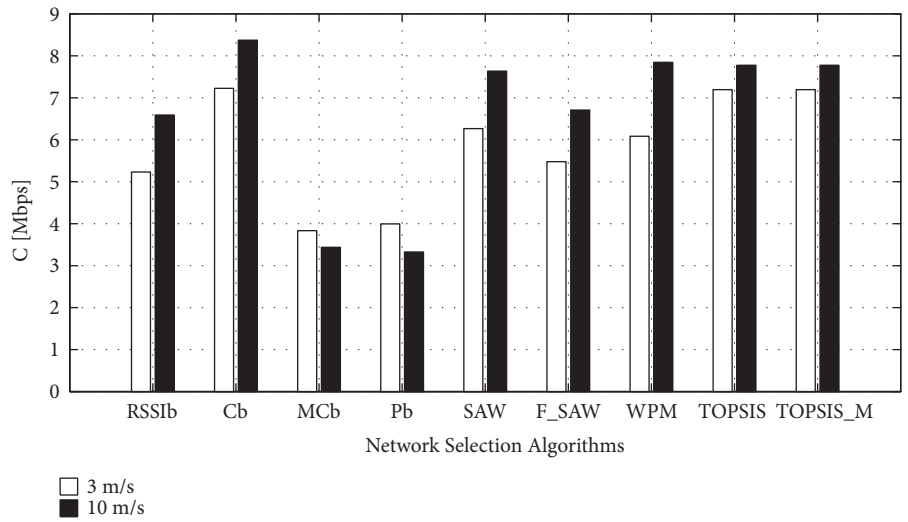


Figure 11: The considered *C* metric for various network selection techniques.

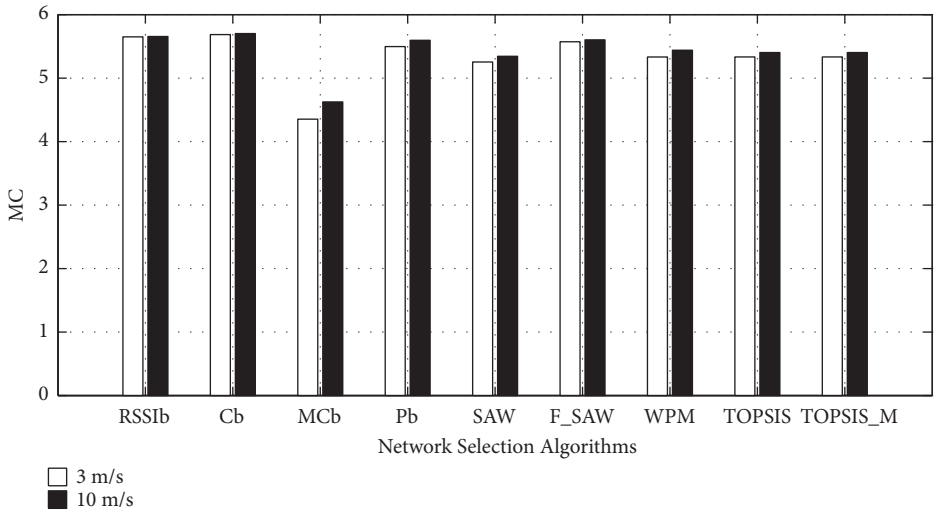


Figure 12: The considered *MC* metric for various network selection techniques.

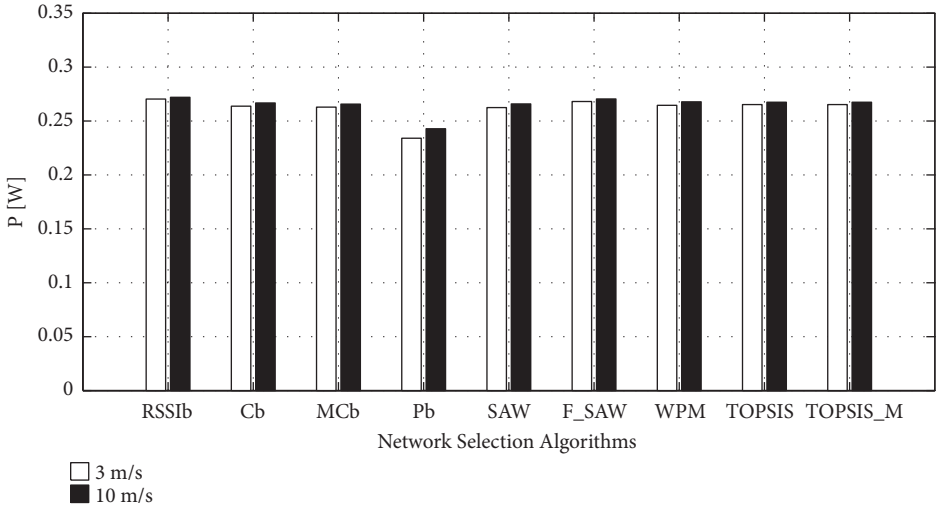


Figure 13: The considered P metric for various network selection techniques.

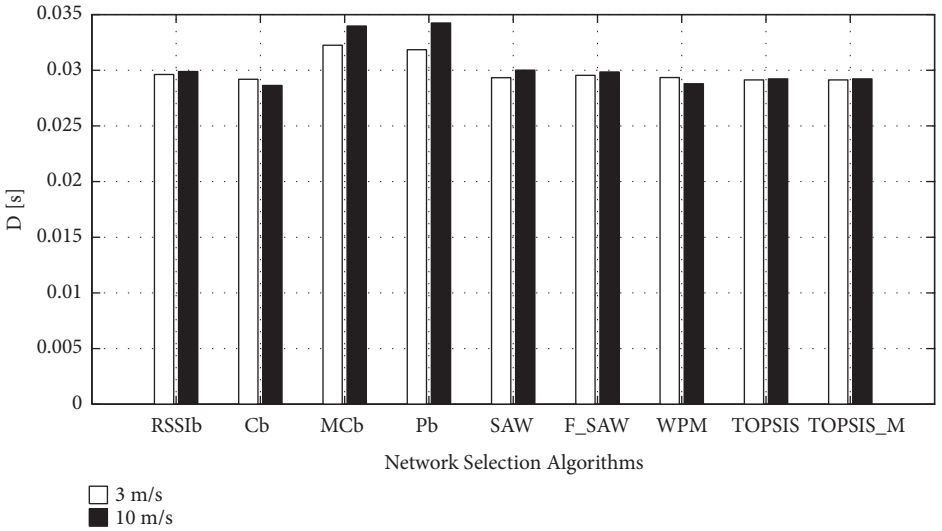


Figure 14: The D metric for various network selection techniques.

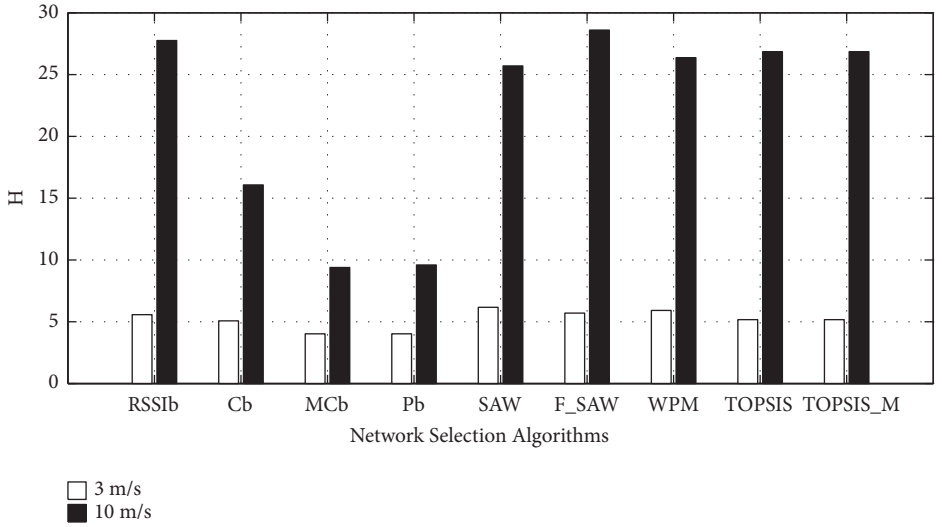


Figure 15: The H metric for various network selection techniques.

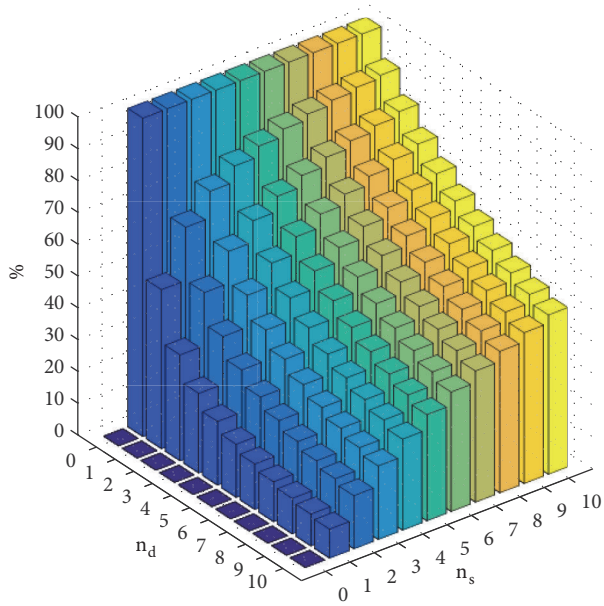


Figure 16: Percentage of reduction in the number of operations of first group using D-TOPSIS respect to TOPSIS.

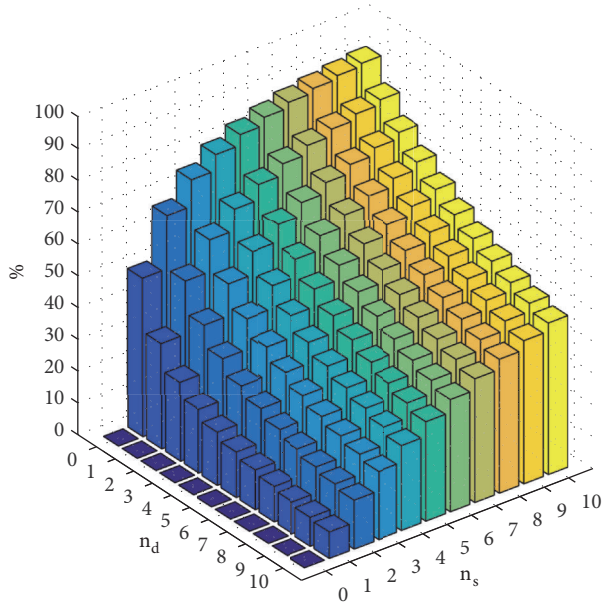


Figure 17: Percentage of reduction in the number of operations of second group using D-TOPSIS compared to TOPSIS.

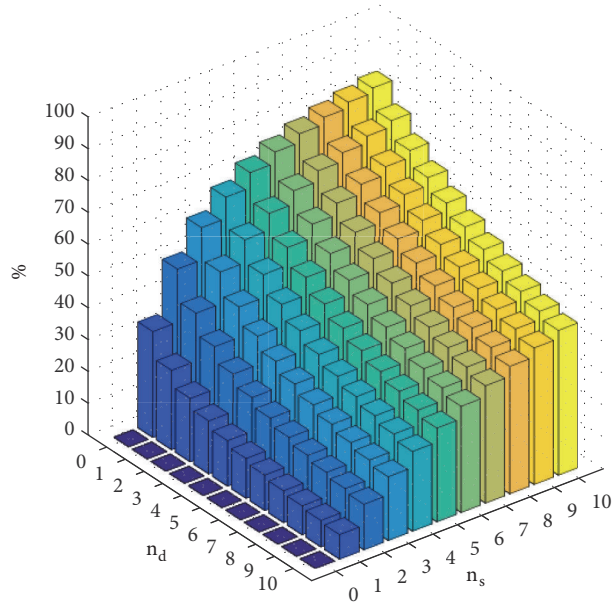


Figure 18: Percentage of reduction in the number of operations of third group using D-TOPSIS compared to TOPSIS.

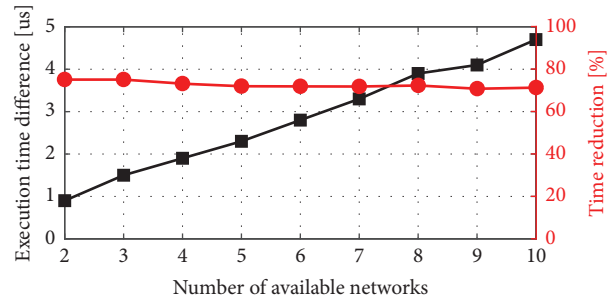


Figure 19: Execution times of TOPSIS and D-TOPSIS and percentage of execution time reduction of D-TOPSIS over different number of *alternatives*.

respect to the standard TOPSIS implementation as highlighted in Section 5. Now the question is, how much does this complexity reduction impact the execution time? So in this subsection a comparison of the execution time of both the TOPSIS implementation versions is proposed. Moreover also the execution time of the algorithms cited in Section 4 is evaluated and compared.

Figure 19 highlights 2 different quantities: the difference in the execution time between the two different implementation versions of the TOPSIS algorithm (see left ordinate axis) and the percentage of the execution time reduction of the D-TOPSIS with respect to the TOPSIS (see right ordinate axis). This quantity is calculated as the difference between the two execution times, divided by the TOPSIS execution time. The total number of attributes is four ($n = 4$) while the number of static and dynamic attributes is, respectively, three and one. In practice, $n_s = 3$ and $n_d = 1$.

From the figure we can extrapolate the idea that the difference between the execution times of the two versions of the TOPSIS algorithm gets bigger if the number of *alternatives* increases. On the other hand, this is not true for the reduction of execution time [31]. Indeed, such quantity is constantly between 70% and 75% independently of the number of *alternatives*. In practice, these results confirm that the execution time reduction is independent of the number of available *alternatives*, as anticipated in Section 5. Figure 20 shows the execution time for all the considered network selection algorithms for a different number of RANs with $n_s = 3$ and $n_d = 1$. One more time, D-TOPSIS is the

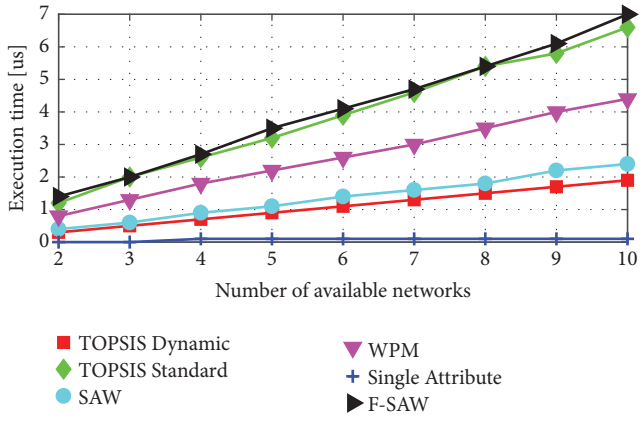


FIGURE 20: Execution time of the considered network selection algorithms over different number of *alternatives*.

second fastest algorithm among the considered algorithms. As a matter of fact the algorithms that belong to this group are computationally lighter with respect to the multiattribute approaches and select the network considering a single parameter applying only an *argmax* or *argmin* operator. On the other hand, these algorithms give poor results and a suboptimal selection [31].

Further considerations can be made observing the multiattribute algorithms: the D-TOPSIS can assure good results, in terms of fast execution with respect to all the considered network selection algorithms. As is shown in Figure 20, the D-TOPSIS is the fastest algorithm among the multiattribute ones, while the standard TOPSIS is set in a middle range and the fuzzy solution requires the maximum execution time. This is due to the fact that it needs many operations during the *fuzzification* and *defuzzification* steps.

Finally, we can observe how the difference between the running times gets bigger when the number of *alternatives* increases. If $m = 10$, the D-TOPSIS running time is equal to 1.9 [μs]. For the TOPSIS standard such quantity is 6.6 [μs], while, for the *Fuzzy Logic*, it has the value of 7 [μs].

7. Conclusions

The paper considers the problem of the network selection when different Radio Access Networks (RANs) that employ different Radio Access Technologies (RATs) are available and can be used to guarantee communications in case of emergency event during which more than one network has to be used to allow a continuous exchange of information between a Mobile Node (MN) and a remote host.

In more detail, in the general framework of the IEEE 802.21 standard, a Decision Maker (DM), within the protocol stack of the MN, in charge of performing network selections and handover decisions, has been proposed. From the mathematical viewpoint, the Multiattribute Decision Making (MADM) techniques are considered and in particular the TOPSIS approach, which has been made lighter (and called D-TOPSIS) with respect to the computational burden, needed to select the RAT. The concept applied is that the *attributes* (i.e., the performance metrics) used to evaluate an *alternative* (i.e., a RAT) can be divided into two groups: the *static attributes* and the *dynamic attributes*. The static attributes for each *alternative* maintain their values constant when the *alternative* is available. The *dynamic attributes* change over time. The basic idea is to run TOPSIS each time the set of available RANs changes (i.e., the MN enters or leaves a new RAN), to find a partial solution (based on the *static attributes*) and reuse it in each successive decision selection. This is D-TOPSIS.

A large simulative campaign, aimed at comparing the performance and the running time of the D-TOPSIS, the TOPSIS, and the technique found in the literature, has been finally reported. MADM approaches guarantee a satisfactorily compromise performance for all the considered metrics, *RSSI*, Capacity (C), Monetary Cost (MC), power consumption (P), packets delay (D), and the number of handover processes executed by the MN (H)). On the other hand, in terms of computation burden, for a single network selection decision, D-TOPSIS allows saving from 10% to 70% of the time needed by the other MADM decision techniques to provide the selected RAT.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors want to deeply thank Dr. Stefano Delucchi. The preparation of this manuscript would not be possible without his precious work and his crucial contribution.

References

[1] M. Frodigh, S. Parkvall, C. Roobol, P. Johansson, and P. Larsson, "Future-generation wireless networks," *IEEE Personal Communications*, vol. 8, no. 5, pp. 10–17, 2001.

[2] "IEEE Standard for Local and metropolitan area networks - Media Independent Handover Services," IEEE Std 802.21-2008, pp. c1–301, 2009.

[3] I. Chantaksinopas, P. Oothongsap, and A. Prayote, "Network selection delay comparison of network selection techniques for safety applications on vanet," *Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific*, pp. 1–7, 2011.

[4] L. Wang and G.-S. Kuo, "Mathematical modeling for network selection in heterogeneous wireless networks: a tutorial," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 1, pp. 271–292, 2013.

[5] P. K. Yoon, C. L. Hwang, and K. Yoon, *Multiple Attribute Decision Making: An Introduction (Quantitative Applications in the Social Sciences)*, SAGE Publishing, 1995.

[6] A. F. Molisch, *Wireless Communications*, Wiley, 2011.

[7] I. Bisio, C. Braccini, S. Delucchi, F. Lavagetto, and M. Marchese, "Dynamic multi-attribute network selection algorithm for vertical handover procedures over mobile ad hoc networks," in *Proceedings of the 1st IEEE International Conference on Communications (ICC)*, pp. 342–347, IEEE, 2014.

[8] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: media-independent handover services," *IEEE Wireless Communications Magazine*, vol. 15, no. 4, pp. 96–103, 2008.

[9] G. Chen, M. Song, Y. Zhang, X. Wang, and B. Sun, "End-to-end QoS guaranteed seamless handover scheme for cognitive heterogeneous networks," in *Proceedings of the 2nd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 440–444, IEEE, 2010.

[10] S. Mohanty and I. F. Akyildiz, "A cross-layer (layer 2 + 3) handoff management protocol for next-generation wireless systems," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1347–1360, 2006.

[11] C. Wang, K. Sohraby, R. Jana, L. Ji, and M. Daneshmand, "Network selection for secondary users in cognitive radio systems," in *Proceedings of the IEEE INFOCOM*, pp. 2741–2749, 2011.

[12] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369–395, 2004.

[13] W. Shen and Q.-A. Zeng, "Cost-function-based network selection strategy in integrated wireless and mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3778–3788, 2008.

[14] H. J. Wang, R. H. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA)*, pp. 51–61, IEEE, 1999.

[15] R. Chai, L. Tang, M. Xiao, and Q.-B. Chen, "Adaptive network selection algorithm based on user application profile," in *Proceedings of the 5th International ICST Conference on Communications and Networking in China (ChinaCom)*, pp. 1–5, 2010.

[16] I. Bisio, S. Delucchi, F. Lavagetto, and M. Marchese, "Capacity bound of MOP-based allocation with packet loss and power metrics in satellite communications systems," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 3311–3316, IEEE, 2012.

[17] H. Liu, C. Maciocco, V. Kesavan, and A. L. Y. Low, "Energy efficient network selection and seamless handovers in mixed networks," in *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks and Workshops (WOWMOM)*, pp. 1–9, IEEE, 2009.

[18] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *Proceedings of the IEEE Network Operations and Management Symposium (NOMS)*, pp. 144–151, IEEE, 2008.

[19] K. Yoon, "The propagation of errors in multiple-attribute decision analysis: a practical approach," *Journal of the Operational Research Society*, vol. 40, no. 7, pp. 681–686, 1989.

[20] H. Yaïche, R. R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.

[21] Q. He, G. Chen, and L. Zhang, "A vertical handoff decision algorithm based on fuzzy control in WiMAX and TD-SCDMA heterogeneous wireless networks," in *Proceedings of the 6th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, pp. 1–4, 2010.

[22] J. Hou and D. C. O'Brien, "Vertical handover decision-making algorithm using fuzzy logic for the integrated radio-and-OW system," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 176–185, 2006.

[23] Q. He, "A fuzzy logic based vertical handoff decision algorithm between WWAN and WLAN," in *Proceedings of the 2nd International Conference on Networking and Digital Society (lCNDS)*, vol. 2, pp. 561–564, 2010.

[24] J. Ding, X. Xue, and G. Li, "Rss method and fuzzy logic combined vertical handoff decision algorithm," in *Proceedings of the IET International Communication Conference on Wireless Mobile & Computing (CCWMC)*, pp. 137–139, Shanghai, China, 2009.

[25] A. Singhrova and N. Prakash, "Adaptive vertical handoff decision algorithm for wireless heterogeneous networks," in *Proceedings of the 11th IEEE International Conference on High Performance Computing and Communications (HPCC)*, pp. 476–481, IEEE, Republic of Korea, 2009.

[26] I. Bisio, A. Delfino, S. Delucchi et al., "Simulative performance comparison of network selection algorithms in the framework of the 802.21 standard," in *Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, IEEE, 2013.

[27] M. Lahby, L. Cherkaoui, and A. Adib, "Network selection algorithm based on Diff-AHP and TOPSIS in heterogeneous wireless networks," in *Proceedings of the International Conference on Multimedia Computing and Systems (ICMCS)*, pp. 485–490, Morocco, 2012.

[28] A. Sgora, C. A. Gizelis, and D. D. Vergados, "Network selection in a WiMAXWiFi environment," *Pervasive and Mobile Computing*, vol. 7, no. 5, pp. 584–594, 2011.

[29] I. Bisio, S. Delucchi, F. Lavagetto, and M. Marchese, "Performance comparison of network selection algorithms in the framework of the 802.21 standard," *Journal of Networks*, vol. 1, pp. 51–59, 2015.

[30] U. D. of Commerce, "Nist national institue of standards and technology," http://www.nist.gov/index.html.

[31] I. Bisio, C. Braccini, S. Delucchi, F. Lavagetto, and M. Marchese, "Performance evaluation of network selection algorithms for vertical handover procedures over satellite/terrestrial mobile networks," in *Proceedings of the 6th International Conference on Advances in Satellite and Space Communications (SPACOMM)*, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7521513, 20 pages
https://doi.org/10.1155/2019/7521513

Research Article

Deploying a Reliable UAV-Aided Communication Service in Disaster Areas

Vicente Mayor, Rafael Estepa, Antonio Estepa, and German Madinabeitia

Department of Telematics Engineering, Escuela Superior de Ingenieros, Universidad de Sevilla, C/ Camino de los Descubrimientos s/n, 41092 Sevilla, Spain

Correspondence should be addressed to Antonio Estepa; aestepa@us.es

Received 1 November 2018; Revised 24 January 2019; Accepted 31 January 2019; Published 8 April 2019

Academic Editor: Song Guo

When telecommunication infrastructure is damaged by natural disasters, creating a network that can handle voice channels can be vital for search and rescue missions. Unmanned Aerial Vehicles (UAV) equipped with WiFi access points could be rapidly deployed to provide wireless coverage to ground users. This WiFi access network can in turn be used to provide a reliable communication service to be used in search and rescue missions. We formulate a new problem for UAVs optimal deployment which considers not only WiFi coverage but also the mac sublayer (i.e., quality of service). Our goal is to dispatch the minimum number of UAVs for provisioning a WiFi network that enables reliable VoIP communications in disaster scenarios. Among valid solutions, we choose the one that minimizes energy expenditure at the user's WiFi interface card in order to extend ground user's smartphone battery life as much as possible. Solutions are found using well-known heuristics such as K-means clusterization and genetic algorithms. Via numerical results, we show that the IEEE 802.11 standard revision has a decisive impact on the number of UAVs required to cover large areas, and that the user's average energy expenditure (attributable to communications) can be reduced by limiting the maximum altitude for drones or by increasing the VoIP speech quality.

1. Introduction

The use of UAVs in natural disasters has become popular in recent years [1, 2]. During the critical first 72 hours, UAVs can be used for tasks such as situational awareness [3], deploying communication systems [4–6], or search and rescue (SAR) missions [7]. In this work, we aim to dispatch UAVs to deploy a reliable communication system that can be used in SAR missions.

There are many aspects to consider in the communication that takes place during SAR missions. In [8] the authors identify the main building blocks (e.g., commanding, surveying, relaying, etc.) and its communication requirements in terms of tolerance to delay, jitter and minimum throughput. Such requirements should be carefully considered when selecting wireless link technologies for either UAV-to-UAV or UAV-to-Infrastructure communication [9]. Similarly, in [10] the authors identify requirements for a flexible, secure, robust, and QoS-aware emergency response communication system, analyzing current wireless technologies (e.g., WiFi, WiMAX, cellular, TETRA, and Satellite) in the light of their capability to carry voice, video, and push-to-talk emergency communications. In the aforementioned works however, the fitness criteria are based on link properties such as throughput, delay or coverage versus the requirements of a single communication. Nevertheless, QoS performance in shared access networks such as WiFi does not only depend on signal coverage but also on the traffic generated/consumed by ground users sharing the medium, and if speech quality degrades, the communication service cannot be used. Unfortunately, this fact is frequently overlooked.

In this work we take a first step into the problem of deploying a UAV-aided WiFi network that can support real-time communications in SAR missions such as the one illustrated in Figure 1. We consider a reference SAR communication service composed of a QoS-guaranteed bidirectional voice channel for each ground user as well as a broadcast audio channel for announcements. Then, we define a new optimization problem for the deployment of a set of UAVs-mounted Access Points (APs) which consists of finding the minimum number of UAVs (and their position) to provide the aforementioned communication service to a set of known

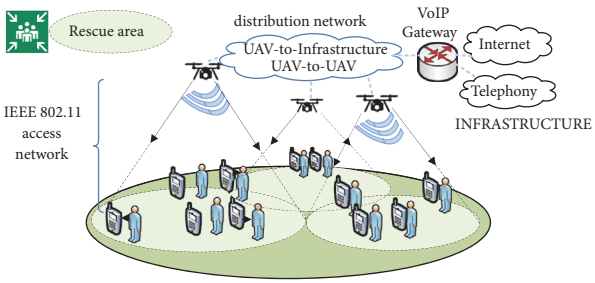


Figure 1: Reliable communication service scenario.

ground targets. Among alternative solutions (i.e., same number of UAVs but in different positions) we choose the one that exhibits the lowest energy expenditure in the network interface card (NIC) of terminals so ground users may extend their battery life. To the best of our knowledge, a similar problem has never been considered before. Note that we assume that the performance of the WiFi access network (see Figure 1) is critical for the feasibility of the QoS-guaranteed service while UAV-to-UAV or UAV-to-Infrastructure does not significantly impair QoS, or its impact can be added to our model as a constant extra delay and/or loss. Therefore, the scope of this paper is restricted to the IEEE 802.11 access network to be deployed.

Our optimization problem could be directly applied to cover known areas (e.g., meeting points or facilities) when the number of users is known or to provide service to semistatic ground users. Besides, we also analyze the limits of the applicability of this problem when ground users are moving.

The contributions and originality of this paper are as follows:

(i) We mathematically formulate a new optimal drone location problem that considers coverage, quality of service, and energy expenditure attributable to the communication.

(ii) We provide a mathematical model to predict speech quality for a set of heterogeneous VoIP traffic sources and for different IEEE 802.11 standard amendments.

(iii) We propose a mathematical model to predict energy expenditure for a set of users and for different IEEE 802.11 standard amendments.

(iv) We provide a preliminary analysis of the applicability of our problem to ground users on the move.

The findings of this paper can be useful not only for deploying a communication network, but also as a first step in the field of media coverage (audio), which is an unexplored area of application in disaster management according to [1].

2. Related Works

The deployment of standalone communication systems in disaster scenarios is not new [1]. UAVs-aided wireless communication systems have been previously proposed to provide ubiquitous coverage (e.g., off-load of existing Base Stations, or functional replacement of damaged ones [4, 9]), relaying by interconnecting groups of distant users [11], or to disseminate or collect information to/from ground terminals [12, 13]. All previous applications can be useful in disaster scenarios. However, the comparison of existing proposals is difficult as each one sets the focus on different aspects of the communication system (e.g., networking schemes and architectures [10, 14, 15], network formation in UAV-to-UAV communication [16], use of hybrid networks with various wireless link technologies [10, 17]) and, generally, does not specify the traffic generated by users, or else such traffic is not voice or video [15].

Most works addressing the optimization of UAV deployment seek to find the minimum number of UAVs required to provide signal coverage (i.e., received signal strength) to a set of ground users at known positions [18, 19] disregarding the traffic generated by ground users. In some cases, the optimization problem seeks to cover all users while minimizing the energy expenditure in the UAV as a result of propulsion [20] or the communication [21].

UAV-mounted IEEE 802.11 Access Points (APs) have also been explored in the past [8, 22–25]. A comprehensive summary of the challenges and implications of the various IEEE 802.11 amendments in long-range outdoor WiFi deployments can be found in [25]. Studies, however, only analyze the performance of IEEE 802.11 links in terms of delay, jitter, or throughput achievable by one traffic flow. While results confirm the technical feasibility of deploying an UAV-aided WiFi network for real-time voice communications, they have a common weakness: only one communication flow is considered, disregarding the negative effect of the traffic generated/consumed by WiFi users in the network performance. Moreover, it is well known [26] that there is an upper bound in the number of simultaneous conversations that each AP can take so the speech quality perceived by VoIP users is acceptable. For this reason, deploying a WiFi network usable in SAR missions should consider this restriction in the number and position of UAVs deployed. However, to the best of our knowledge, no previous work has taken this into consideration before.

3. Guaranteed QoS VoIP Service and Energy Expenditure in WiFi Networks

Speech quality perceived by users largely depends on end-to-end delay and packet loss experienced by VoIP packets, which is related to the performance of the WiFi network. This relation has been investigated over the last decades in its own research field [26–28] but its findings have been seldom applied to AP positioning. This section elaborates on the implications of WiFi network performance in both speech quality and NIC energy expenditure experienced by ground users in our context.

Speech quality can be estimated at the planning stage by using the E-Model [29], which provides a quality score termed R factor (0-100). It is generally agreed that conversations are acceptable when its R factor is above a threshold ($R_{\min}$), commonly $R_{min} = 65$. In its simplest form, R can be expressed as [30]

$$R = 94.2 - I_d - I_{e,eff} \quad (1)$$

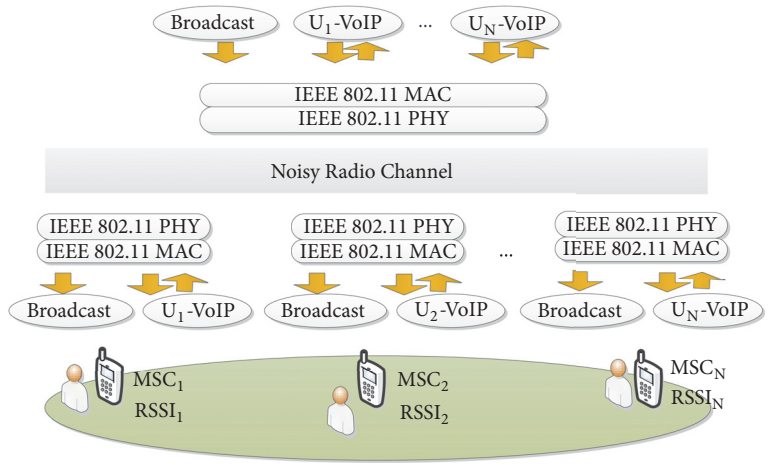


Figure 2: Traffic in the VoWiFi system.

Table 1: 802.11 sensitivity relations for OFDM modulations with 20 MHz channels.

Modulation	Coding rate	Data rate (Mb/s)	Sensitivity (dBm)
BPSK	1/2	6	-82
BPSK	3/4	9	-81
QPSK	1/2	12	-79
QPSK	3/4	18	-77
16-QAM	1/2	24	-74
16-QAM	3/4	36	-70
64-QAM	2/3	48	-66
64-QAM	3/4	54	-65

where, for now, it suffices to know that I_d represents all impairments due to delay and $I_{e,eff}$ is a factor that accounts for the impairments caused by low bit-rate coding and packet loss [31]. Both delay and packet loss are certainly affected by the performance of the IEEE 802.11 access network.

Figure 2 illustrates a system composed of a set of IEEE 802.11 stations associated with one UAV-mounted AP. Each station is represented by its physical layer and its medium access control (MAC) sublayer. While all stations share a common MAC protocol, each user station auto-configure its Modulation and Coding Schema (MCS) according to the Received Signal Strength Indicator (RSSI) from the AP. An example of different MCSs defined for OFDM modulations (e.g., 802.11a/g) is shown in Table 1. Thus, depending on the received signal power, each user terminal will auto-configure its MCS to the greatest bit-rate possible.

At the application level, this work assumes that each ground user handles two real-time VoIP communication channels: one downlink broadcast audio channel to listen to announces from the rescue team and one bidirectional channel for a regular conversation. We assume that the broadcast traffic is always sent at the lowest bit-rate possible.

In this scenario, it is well known that VoIP speech quality is mainly determined by the delay and loss experienced at the AP's downlink transmission buffer [26], which in turn depends on the speech codec used by the VoIP application and the number of simultaneous calls at a given moment (i.e., traffic load). Some studies quantify this effect and determine the maximum number of conversations that can be simultaneously held so that delay and/or loss in the AP does not exceed a certain threshold. For example, in [26] it was found that a maximum of 15 calls for 64kb/s CBR traffic, or 38 calls for VBR VoIP traffic, can be held for IEEE 802.11b.

Thus, UAV placement will be strongly influenced by the maximum number of simultaneous calls that each AP can take so that minimum level of speech quality is met. However, finding this number is a complex task which falls into the WiFi QoS research field. There is a plethora of analytical models of the IEEE 802.11 mac behaviour [32–35]. In most models, the central variable is the probability that an observed station attempts to transmit in a random time slot (τ). However, different models apply different assumptions to derive their analytical expression for τ. In this paper, the following assumptions are made:

(i) Heterogeneous traffic sources (users may have different physical data bit-rate according to their received signal strength).

(ii) Nonsaturated stations (transmission buffer could be empty).

(iii) Noisy channel (packets can be corrupted due to channel noise).

Regarding the energy consumption attributable to the wireless card of IEEE 802.11 stations, one could break it down into the time spent on each possible power state (e.g., transmission, reception, idle) as

$$E = \rho_{tx} \cdot T_{tx} + \rho_{rx} \cdot T_{rx} + \rho_{idle} \cdot T_{idle} \quad (2)$$

where T_{tx}, T_{rx}, and T_{idle} stand for the time spent by the network interface on transmission, reception, and idle states and ρ_{tx}, ρ_{rx}, and ρ_{idle} stand for the power consumption of network interface during those states (the WiFi standard offers the possibility of using Power Saving Mode (PSM), which would add a new factor $\rho_{sleep} \cdot \rho_{sleep}$, namely, sleep mode. Nevertheless, time traffic such as VoIP rarely uses the original PSM mode because it can introduce extra delays, so in this work it will not be considered.). According to [36], the average NIC consumption attributable to the VoIP application is 0.15W.

Clearly, aspects like retransmissions due to noise or collisions, or the MCS used by a station have a great influence in T_{tx}, T_{rx}, and T_{idle}. For instance, the station's data bit-rate (which depends on its MCS) will determine the duration of the transmission or reception of a packet. So we can conclude that the distance between drones and users (the greater distance the lower data rate) and the traffic supported by the WiFi network (which again depends on τ) will influence the NIC energy consumption.

Then, the following two aspects are distinctive of our UAV placement problem:

(i) Each AP can only cover as many users as a minimum guaranteed speech quality level allows to (i.e., $R > R_{\min}$).

(ii) From solutions with the same number of UAVs, the one that exhibits the lowest average power consumption will be preferred.

4. Problem Statement

In the scenario illustrated in Figure 1, we want to deploy a set of drones to create a WiFi access network that can be used to provide the service described in Section 1. The following assumptions are made for the sake of tractability:

(i) The position of ground targets (i.e., users confined in a known area) is known.

(ii) Users have a smartphone running an app which handles the VoIP communications described previously using a known codec.

(iii) Channelisation between APs is done in such a manner that interferences are negligible.

4.1. Terminology. We discretize the flying zone as illustrated in Figure 3. The set $\mathscr{P}$ represents the set of 3D coordinates of each edge of the grid (i.e., potential locations of a drone). The following terms and definitions will be used for the remainder of this paper:

(i) Users are denoted by the set $\mathscr{U} = \{1, 2, \ldots, U\}$ and at known locations given by $\{\boldsymbol{w}_k \mid k \in \mathscr{U}\}$, where $\boldsymbol{w}_k \in \mathbb{R}^3$ represents the 3D coordinates of user k. There are $U = 12$ ground users illustrated in Figure 3.

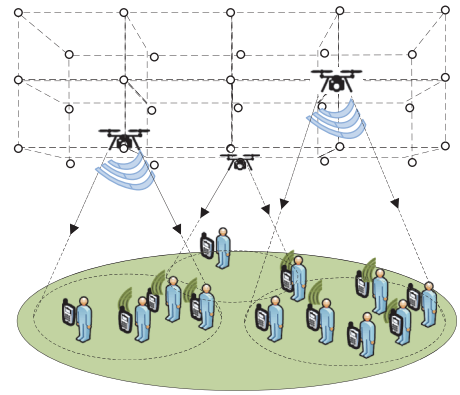


Figure 3: UAVs placement scenario.

(ii) Drones are denoted by the set $\mathscr{D} = \{1, 2, \ldots, D\}$ and at known locations given by the set $\mathscr{X} = \{\boldsymbol{x}_i \mid i \in \mathscr{D}, \boldsymbol{x}_i \in \mathscr{P}\}$, where $\boldsymbol{x}_i$ represents the 3D coordinates of drone i. There are $D = 3$ drones in Figure 3.

(iii) $\mathscr{C}(i) \subset \mathscr{U}$ represents the set of users associated with the AP installed at drone $i \in \mathscr{D}$. Figure 3 illustrates the sets $\mathscr{C}(1)$, $\mathscr{C}(2)$, and $\mathscr{C}(3)$ with 4, 1, and 7 users, respectively.

(iv) Let δ_{ij} be 1 if user $i \in \mathscr{C}(j)$ and 0 otherwise. Then, $C = \sum_{j=1}^{D} \sum_{i=1}^{U} \delta_{ij} = \sum_{i=1}^{D} |\mathscr{C}(i)|$ is the number of ground users associated with the WiFi network ($C = 12 = U$ in Figure 3).

(v) $R(i)$ represents the speech quality level for the users associated with the AP at drone $i \in \mathscr{D}$.

(vi) E_{ij} represents the average energy consumption per second of the WiFi network card installed at user i's terminal when it is associated with drone $j \in \mathscr{D}$.

(vii) $\overline{E}(j)$ represents the NIC average energy consumption of stations associated with the AP at drone j. Then $\overline{E}(j) = \sum_{i=1}^{U} E_{ij}\delta_{ij}/|\mathscr{C}(j)|$.

4.2. Problem Definition. Our goal is to minimize the number of drones deployed to provide service to ground users and, among solutions with the same number of drones, minimize the NICs average power consumption attributable to VoWiFi. This can be formulated as

$$\begin{aligned} \min_{\mathscr{X}} \quad & D + \frac{1}{D} \sum_{i=1}^{D} \frac{\overline{E}(i)}{E_{\max}} \\ \text{subject to} \quad & \sum_{j=1}^{D} \sum_{i=1}^{U} \delta_{ij} = U \\ & R(i) \geq \mathrm{R}_{\min}, \quad \forall i \in \mathscr{D} \\ & \boldsymbol{x}_i \in \mathscr{P}, \quad \forall i \in \mathscr{D} \\ & D \leq D_{\max} \end{aligned} \quad (3)$$

where D represents the number of drones (which is bounded by $D_{\max}$), $E_{\max}$ is a constant that represents the

maximum NIC energy consumption possible (i.e., a station in the highest consumption state, e.g., transmission), and $\overline{E}(i)$ is the average NIC energy consumption of stations associated with drone i. The constant $R_{\min}$ represents the minimum acceptable speech quality (e.g., 65). Observe that the objective function is composed of a principal part (number of UAVs, which is an integer between 1 and $D_{\max}$, and a subordinated one (which is < 1), so its integer part always indicates the minimum number of UAVs that meet all constraints.

By minimizing the number of drones launched, the deployment cost is reduced. Besides, choosing the configuration that minimizes the average energy expenditure in stations also benefits the SAR mission as it prolongs the battery life of ground users' terminals.

5. Solving by Exhaustive Search

The objective function can be evaluated for all edges of the grid $\mathscr{P}$ with one drone ($D = 1$), if no solution is found, the number of drones will be increased ($D = 2$) and every possible combination of positions $\{\boldsymbol{x}_1, \boldsymbol{x}_2 \mid \boldsymbol{x}_1 \in \mathscr{P}, \boldsymbol{x}_2 \in \mathscr{P}, \boldsymbol{x}_1 \neq \boldsymbol{x}_2\}$ will be evaluated, and so on, until an optimal solution is found or a maximum number of UAVs is reached ($D_{\max}$).

Algorithm 1 indicates the main steps followed. The input of the algorithm is the set of ground users and their location, the set of edges $\mathscr{P}$, $D_{\max}$, and the level of speech quality required $R_{\min}$. For an incremental number of drones (D), all possible combinations of UAVs locations $\mathscr{X}$ are checked. Firstly, the number of users associated with each drone in the WiFi network is evaluated. If the first constraint is met, the QoS constraint is assessed (considering the lowest R among all UAVs). If both constraints are met, a potential solution is found and the objective function is evaluated. Among potential solutions with the same value of D, the one with the lowest objective function is selected. The algorithm ends after finding the optimal solution, or after trying unsuccessfully with $D_{\max}$ drones. The output is the set of optimal UAVs locations and the value of the objective function. A null location will be returned if no solution is found.

Next, we elaborate on the functions used for the assessment of signal coverage, speech quality, and power consumption.

5.1. Signal Coverage Evaluation: Associate(). A WiFi station requires Signal to Noise Ratio (SNR) and RSSI to be above minimum levels to properly demodulate the signal from an AP. In case of multiple APs (i.e., $D > 1$), a station associates to the one with greatest RSSI.

Let $RSSI_{ij}$ be the power received by user i located at $\boldsymbol{w}_i$ from the UAV j located at $\boldsymbol{x}_j$. We assume an open area and Line of Sight. Although more complex path loss models could be used (e.g., [37]) we use a free space propagation model:

$$\mathrm{RSSI}_{ij} = P_{TX} + 147{,}55 + G_{ij} - n \cdot 10\log_{10}\left(\left\|\boldsymbol{w}_i - \boldsymbol{x}_j\right\|\right) - 20\log_{10}(f) \tag{4}$$

$$G_{ij} = 10\log_{10}\left(10^{G_{\max}/20} \cdot \cos^2\theta_{ij}\right) \tag{5}$$

Input: $\mathscr{U}$, $\{w_k\}$, $\mathscr{P}$, $D_{\max}$, $R_{\min}$
Output: $\mathscr{L}$ (location), $of_{\min}$ (obj. function)
Initialization: $D = 0$, $\mathscr{L} = \varnothing$, $of_{\min} = D_{\max} + 1$

(1) **while** ($\mathscr{L} = \varnothing$ ***or*** $D \leq D_{\max}$) **do**
(2) D++; // `increase UAVs`
(3) **for** $x \in \mathscr{X}$ **do**
(4) **for** $j = 1$ **to** D **do**
(5) $C(j)$ = associate($\mathscr{U}$,$\{w_k\}$,$\mathscr{X}$);
(6) **end**
(7) **if** $\sum_{i=1}^{D} |\mathscr{C}(i)| == U$ **then** /* `first constaint met` */
(8) **for** $j = 1$ **to** D **do**
(9) $R(j)$=QoSEval($\mathscr{C}(j)$,x_j);
(10) $\overline{E}(j)$=EnergyEval($\mathscr{C}(j)$,x_j);
(11) **end**
(12) $R = \min_{k=1..D}\{\mathrm{R}(k)\}$;
(13) **if** $R \geq R_{\min}$ **then** /* `candidate solution` */
(14) $\mathrm{of} = D + \frac{1}{D}\sum_{i=1}^{D}\frac{\overline{E}(i)}{E_{\max}}$;
(15) **if** $of < of_{\min}$ **then**
(16) $\mathrm{of}_{\min} = \mathrm{of}$;
(17) $\mathscr{L} = \mathscr{X}$;
(18) **end**
(19) **end**
(20) **end**
(21) **end**
(22) **end**

Algorithm 1: Exhaustive search algorithm.

where P_{TX} (dBm) stands for the power delivered by the transmitter antennas; G_{ij} represents the gain of the antenna between user i and UAV j as indicated in (5); n is the propagation exponent (e.g., approximately 3.3 in outdoor); $\|\boldsymbol{w}_i - \boldsymbol{x}_j\|$ is the Euclidean distance between user i and drone j; f (hertz) is the channel frequency. In (5), θ_{ij} (radians) accounts for the elevation angle between user's position $\boldsymbol{w}_i$ and UAV's position $\boldsymbol{x}_j$ and $G_{\max}$ is a constant indicating the maximum power.

Given RSSI_{ij}, the *Signal to Noise Ratio* (SNR_{ij})(dB) can be readily obtained by subtracting the receiver's noise figure (NF) and thermal noise (N) as indicated in (6) and (7).

$$\mathrm{SNR}_{ij} = \mathrm{RSSI}_{ij} - NF - N \tag{6}$$

$$N = -174 + 10\log_{10}(C_{BW}) \tag{7}$$

where C_{BW} is signal bandwidth as specified by the IEEE 802.11 standard in use.

For all $i \in \mathscr{U}$ and for all $j \in \mathscr{D}$, let us define γ_{ij} as a Boolean variable that represents whether user i satisfies minimum thresholds $\mathrm{SNR}_{ij} \geq \mathrm{SNR}_{\min}$ and $\mathrm{RSSI}_{ij} \geq \mathrm{RSSI}_{\min}$ with respect to UAV j or not. Then, the function *associate()* returns, for a specific UAV k, the set of users that meet

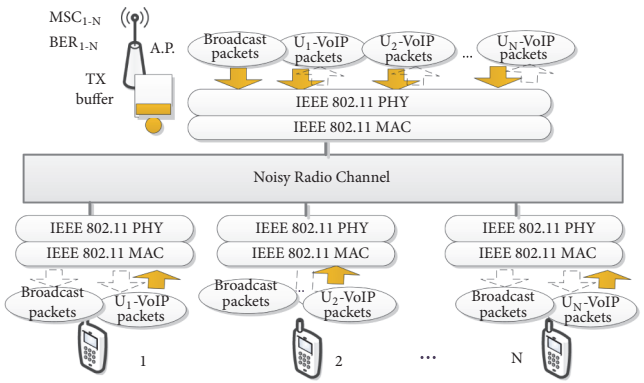


Figure 4: WiFi system in a generic set $\mathscr{C}(k)$.

minimum thresholds and whose RSSI with drone k is greater than with any other drone:

$$\mathscr{C}(k) = \left\{ i \in \mathscr{U} \mid \gamma_{ik} = 1,\ \mathrm{RSSI}_{ik} > \mathrm{RSSI}_{ij}\ \forall j \neq k \right\} \tag{8}$$

5.2. VoWiFi Service QoS Evaluation: QoS.eval(). This function uses an analytical model to estimate the delay and packet loss experienced by the set of users associated with a specific AP. This is then used as input to the E-Model to assess the speech quality level of the set of users associated with a particular UAV.

Remember from Section 3 that each user will be executing an app which handles a broadcast VoIP traffic flow (for announcements) and a bidirectional VoIP conversation. Traffic from all users will be aggregated in the AP-mounted UAV to which users are associated which will act as transparent bridge. Each station is autoconfigured with a specific MCS which determines its data bit-rate and bit error rate (BER). This is illustrated in Figure 4.

All stations are assumed to run the same *Distributed Coordination Function* (DCF) of the 802.11 MAC sublayer, which uses CSMA/CA (*Carrier Sense Multiple Access with Collision Avoidance*) for medium access control. In summary CSMA/CA works as follows: each contending station must sense the medium during a period of time in order to ensure that it is idle before transmission. If the channel is busy, the station waits a random backoff interval before trying again. The backoff process is based on the *Binary Exponential* algorithm. Time is discretized by defining a time slot duration, and the algorithm picks a random number of time slots between 0 and $2^i W_o$, where W_o accounts for the minimum contention window value, and i increases by one in each failed attempt up to a ceiling m. After a maximum number of retransmissions is reached (M), the packet is discarded.

5.2.1. Delay and Packet Loss Estimation. The set of stations in $\mathscr{C}(k)$ and its AP (see Figure 4) constitute a system whose performance has been largely studied in scientific literature [38–41]. We have used a mix of different analytical models from [32–35] suited for the conditions assumed in Section 3: heterogeneous traffic sources, stations are nonsaturated, and noisy channel. For the sake of clarity, we keep this section as simple as possible using a homogeneous notation. However, the reader is encouraged to read more elaborated information in the corresponding references.

The probability that an observed station attempts to transmit in a randomly chosen time slot (τ) can be modelled as in [32]:

$$\tau = \frac{1}{\eta}\frac{1}{1-q}\left(\frac{r^2 W_o}{(1-p)\left(1-(1-r)^{W_o}\right)} - qr(1-p)\right) \tag{9}$$

$$\begin{aligned}\eta = {} & (1-r) + \frac{r^2 W_o (W_o+1)}{2\left(1-(1-r)^{W_o}\right)} \\ & + \frac{W_o+1}{2(1-q)}\left(\frac{r^2 q W_o}{1-(1-r)^{W_o}} + rp(1-q)\right. \\ & \left. - rq(1-p)^2\right) + \frac{p}{2(1-q)(1-p)}\left(\frac{r^2 W_o}{1-(1-r)^{W_o}}\right. \\ & \left. + qr(1-p)^2\right)\left(2W_o\frac{1-p-p(2p)^{m-1}}{1-2p}+1\right)\end{aligned} \tag{10}$$

where η is defined in (10), q is the probability of having at least one packet queued at the transmission buffer after an average MAC service time, p is the probability that a packet suffers any transmission errors, and r is the probability that at least one packet arrives during an idle state. Assuming nonsaturated stations and Poisson packet arrivals to the queue (with rate λ), r and q can be expressed as

$$r = 1 - e^{-\lambda E[T]} \tag{11}$$

$$q = 1 - e^{-\lambda E[T] E[B]} \tag{12}$$

where $E[T]$ represents the expected average slot duration and $E[B]$ is expected average number of backoff slots that a packet waits before transmission. Due to its complexity, a closed-form of $E[T]$ is deduced in Appendix A. $E[B]$ can, however, be expressed as in [35]:

$$E[B] = \frac{W_o}{2(1-p)}\left(\frac{1-p-(2p)^m}{(1-2p)} - 2^m p^{M+1}\right) \tag{13}$$

Let $\tau^{(j)}$, $r^{(j)}$, $q^{(j)}$, $E[B^{(j)}]$ and $\lambda^{(j)}$ denote τ, r, q, $E[B]$ and λ for the station j in the system under consideration (i.e., ground users and AP such as in Figure 4), where j refers to either one user station $j = \{1, 2, .. |\mathscr{C}(k)|\}$, or the AP ($j = |\mathscr{C}(k)| + 1 = AP$). Let $p^{(j)}$ be the probability of packet transmission error p for station j. Then, $p^{(j)}$ can be broken down as

$$p^{(j)} = \left(1 - P_i^{(j)}\right) + \mathrm{FER}^{(j)} - \left(1 - P_i^{(j)}\right) \cdot \mathrm{FER}^{(j)} \tag{14}$$

where $P_i^{(j)}$ accounts for the probability that the *j-th* station finds the channel idle and $\mathrm{FER}^{(j)}$ stands for the Frame Error

Rate due to channel noise. The probability of finding the channel idle can be further expressed as

$$P_i^{(j)} = \prod_{i=1, i\neq j}^{S} \left(1 - \tau^{(i)}\right) \tag{15}$$

where S stands for the number of stations in the system, including the AP (i.e., $|\mathscr{C}(k)| + 1$).

Assuming that frames have a constant size of L bits (L = preamble + header + data), it is possible to obtain the FER of station j as

$$\text{FER}^{(j)} = 1 - \left(1 - \text{BER}^{(j)}\right)^{L} \tag{16}$$

where $\text{BER}^{(j)}$ represents the bit error rate of station j, which can be readily calculated (see BER equations in [42] and [43] for DSS and OFDM modulations respectively) if one knows user's modulation (i.e., MCS_j).

Finally, solving the nonlinear equation system, the *packet loss* of station j can be expressed as

$$\text{PL}^{(j)} = 1 - \frac{\left(1 - \text{FER}^{(j)}\right)\tau^{(j)}\prod_{i=1, i\neq j}^{S}\left(1 - \tau^{(i)}\right)}{\lambda^{(j)} E\left[T\right]} \tag{17}$$

Since we assume a very small buffer size, the queueing delay can be neglected and the only delay component will be the channel access delay. Then, the delay of a station j can be expressed as

$$\text{DEL}^{(j)} = E\left[B^{(j)}\right] E\left[T\right] \tag{18}$$

As justified in [26], the AP ($j = AP$) is the most saturated station, leading the packet loss and delay in the system (observe that the AP is also part of the system and its MCS and FER change dynamically according to its communication partner. Our approach is to consider average values of its data bit-rate and FER.). Thus, we can take its packet loss and delay as representative of the worst case. Then, the output of this step is

$$\text{PL} = \text{PL}^{(AP)} \tag{19}$$

$$\text{DEL} = \text{DEL}^{(AP)} \tag{20}$$

5.2.2. Speech Quality Estimation. As stated in Section 3, the E-Model rates the conversation quality R factor, which can be calculated using (1) [44], whose terms were

(i) $I_{e,eff}$ is the effective impairment equipment parameter, which is a combination between the impairment equipment parameter at zero packet loss (I_e), and a function of I_e that is dependent on packet loss rate and packet loss behaviour. It can be expressed as

$$I_{e,eff} = I_e + \left(95 - I_e\right)\frac{P_{pl}}{P_{pl}/\text{BurstR} + B_{pl}} \tag{21}$$

where I_e is a codec-dependent constant associated with codec compression degradation (a list of values from ITU-T codecs were presented in ITUT-T Rec. G.113 Appendix I), P_{pl} represents the packet loss rate, BurstR accounts for the burst ratio (i.e., equals 1 if packet loss if random and greater otherwise), and B_{pl} represents the codec packet loss robustness, which also has a specific value for each codec (listed in ITU-T Rec. G117 Appendix I).

(ii) I_d accounts for all impairments due to delay of communication chain. A widely accepted approximation for I_d can be obtained from one-way delay in communication path (d) as follows:

$$I_d = 0.024d + 0.11\left(d - 177.3\right)H\left(d - 177.3\right) \tag{22}$$

where H is the heavy side function (i.e., $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x > 0$). This shows that in practical terms, small delays (e.g. <100 ms) can be disregarded.

In this paper we will use the G.711 codec ($I_e = 0$), with BurstR = 1 and $B_{pl} = 25.1$. So, assuming that $P_{pl} = \text{PL}$ (from Eq. (19)) and $d = \text{DEL}$ (from Eq. (20)) + 20 ms (from the VoIP codec packetization), R for the set of users associated with a drone can be expressed (the 20 ms of packetization delay are included) as

$$\begin{aligned} R = 93.72 - 95\frac{\text{PL}}{\text{PL} + 25.1} - [0.024\text{DEL} \\ + 0.11\left(\text{DEL} - 157.3\right)H\left(\text{DEL} - 157.3\right)] \end{aligned} \tag{23}$$

5.3. Energy at Terminals: EnergyEval(). As stated in Section 3 the energy consumption depends on how long the NIC spends on each one of the possible energy states (TX, RX or IDLE) as shown in (2). Considering all disjoint events that can happen during a slot, the energy consumption at the NIC of station j can be further decomposed as

$$\begin{aligned} E\left[J^{(j)}\right] = {} & \underbrace{J_\sigma^{(j)} P_i}_{\text{idle interval}} + \underbrace{J_{tx,s}^{(j)} P_{tx,s}^{(j)} + J_{tx,e}^{(j)} P_{tx,e}^{(j)} + J_{tx,c}^{(j)} P_{tx,c}^{(j)}}_{\text{tranmission interval}} \\ & + \underbrace{J_{rx,s}^{(j)} P_{rx,s}^{(j)} + J_{rx,e}^{(j)} P_{rx,e}^{(j)} + J_{rx,c}^{(j)} P_{rx,c}^{(j)}}_{\text{reception interval}} \\ & + \underbrace{J_{o,s}^{(j)} P_{o,s}^{(j)}}_{\text{other stations}} \end{aligned} \tag{24}$$

where $J_{tx,s}^{(j)}$, $J_{tx,e}^{(j)}$, and $J_{tx,c}^{(j)}$ represent the expected energy consumption during a successful, erroneous, or collided transmission, respectively, $J_{rx,s}^{(j)}$, $J_{rx,e}^{(j)}$, and $J_{rx,c}^{(j)}$ represent the expected energy consumption during a successful, erroneous, or collided reception respectively, and $J_{o,s}^{(j)}$ represents the expected energy consumption when listening to a packet successfully sent to other station. The terms $P_{tx,s}^{(j)}$, $P_{tx,e}^{(j)}$, $P_{tx,c}^{(j)}$, $P_{rx,s}^{(j)}$, $P_{rx,e}^{(j)}$, $P_{rx,c}^{(j)}$, $P_{o,s}^{(j)}$ indicate the probability of such events during a time slot ($E[T]$). Finally, $J_\sigma^{(j)}$ represents the energy spent during an idle interval (i.e., none of the other events happened) and P_i is the probability of that.

As an example, let us analyze the first factor of (24): $J_\sigma^{(j)}$ and P_i. The expected energy consumption during an idle interval of duration σ will be

$$J_\sigma = \rho_{idle} \cdot \sigma \tag{25}$$

Table 2: Solution obtained versus simulation.

Drone (i)	x_i	$C(i)$	Algorithm / Simulation	
			$R(i)$ [1 - 100]	$\frac{\overline{E}(i)}{E_{max}}$
1	(20, 20, 20)	11	86 / 87	0.10 / 0.09
2	(30, 70, 20)	11	85 / 87	0.11 / 0.10
3	(80, 30, 20)	15	80 / 83	0.12 / 0.11
4	(80, 80, 20)	13	83 / 85	0.11 / 0.10

Table 3: Example solution input parameters.

IEEE Standard		Scenario		Traffic		Constraints	
Revision	802.11n	Users	50	Call length	180 s	$RSSI_{min}$	-82 dBm
GI	800 ns	Size	100 m × 100 m	VoIP codec	G.711	SNR_{min}	20 dB
Preamble	Greenfield	X-Y step	10 m	On/Off times	CBR	R_{min}	65
Bandwidth	20 MHz	Altitude layers	{20, 30} m	Packet interval	20 ms		
Retries (M)	7	Prop. Exponent	3.3	Broadcast channels	1		

and the probability that stations do not transmit in an interval is

$$P_i = \prod_{j=1}^{S} \left(1 - \tau^{(j)}\right) \tag{26}$$

A detailed expression of the remainder factors in (24) is deduced in Appendix B.

Given that $E[J^{(j)}]$ is the expected energy consumption for each slot, it is possible to carry out the average energy consumption per second of station j as

$$E_j = \frac{E\left[J^{(j)}\right]}{E\left[T\right]} \tag{27}$$

Then, the average energy consumption per second of all ground user stations that belong to $\mathscr{C}(k)$ will be

$$\overline{E} = \frac{\sum_{i \in \mathscr{C}(k)} E_i}{|\mathscr{C}(k)|} \tag{28}$$

which is the output of this function.

6. Example Solution

This section introduces a first example of the results obtained after implementing Algorithm 1 in Matlab®. The scenario includes 50 users randomly distributed among an area of 100 m × 100 m. The X-Y step used was 10 m. The parameters used are listed in Table 3 . The consumption values for ρ_{tx}, ρ_{rx}, and ρ_{idle} were 2.5, 0.9, and 0.11, respectively. Unless otherwise specified, such parameters are common to all experiments in this paper.

Figure 5 provides a graphical representation of the solution obtained, showing users' distribution among UAVs by color. In this case, 4 drones were necessary to service ground users with the standard IEEE 802.11n.

The solution obtained with our exhaustive search algorithm has been used as input scenario to the network simulator ns-3 in order to validate the IEEE 802.11 analytical model provided in Section 5.2. The ns-3 model *YansWifiPhy* with a transmission buffer size of one packet has been used in simulation. Table 2 summarizes the results obtained, comparing R and the normalized energy expenditure from our analytical models with that of the simulation. Results presented are the average of 30 runs.

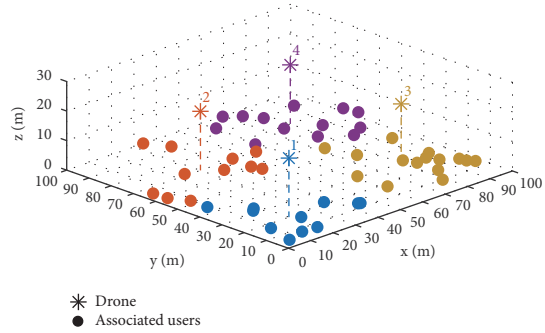


Figure 5: Solution obtained.

Looking at Table 2 one can see acceptable differences in the output of the E-Model, which suggests that the methods and models proposed in Section 5.2 are accurate enough.

While exhaustive search always finds the optimal solution, it exhibits poor scalability since assessing D drones requires to evaluate $\sum_{d=1}^{D} \binom{|\mathscr{P}|}{d}$ possibilities (e.g., $\sim 7 \cdot 10^7$ points in the scenario under consideration). For that reason, exhaustive search is not always feasible and heuristic search methods are commonly used as a general way to find a (semi) optimal solution.

7. Heuristic Solutions

Genetic algorithms (GAs) have demonstrated considerable success in providing good solutions to a wide variety of optimization problems, including UAVs deployment [45, 46]. Other metaheuristic methods, including particle

Input: $\mathcal{U}$, $\{w_k\}$, $\mathcal{P}$, $D_{\max}$, $R_{\min}$
Output: $\mathcal{L}$ (location), $of_{\min}$ (obj. function)
Initialization: $D = 1$, $\mathcal{L} = \varnothing$
(1) **while** ($\mathcal{L} = \varnothing$ *or* $D \leq D_{\max}$) **do**
(2) $\mathcal{L}$, of = *GAsearch*(D, $\{\boldsymbol{w}_k\}$, $R_{\min}$);
/* call GA */
(3) D++;
(4) **end**

Algorithm 2: Heuristic search pseudocode.

swarms optimization, artificial immune system, and simulated annealing, can be used instead or in hybridization with GAs. However, we opt for using GAs for the sake of simplicity since the performance obtained (which is analyzed in the following subsection) is fair enough for the purpose of providing a numerical analysis of our new problem which is our main contribution.

We propose a heuristic search method that replaces lines (2)-(21) in Algorithm 1 by a function call (*GAsearch*) that runs a genetic algorithm to find the optimal solution for a specific number of drones D as shown in Algorithm 2. D is increased until a valid solution is found up to a maximum value of $D_{\max}$ drones.

7.1. Genetic Algorithm: GAsearch(). We have used the Matlab® R2017A Global Optimization Toolbox, whose main operators are defined in [47]. The basic concepts and steps followed are detailed bellow.

7.1.1. Individuals. An individual is defined as a possible solution to the problem. Thus, each individual is a location of D drones $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_D \mid \boldsymbol{x}_i \in \mathcal{P},\ i \in \mathcal{D},\ \boldsymbol{x}_i \neq \boldsymbol{x}_j|_{\forall j \neq i}\}$, where $\boldsymbol{x}_i \in \mathcal{P}$ represents the 3D coordinates of drone i. For example, for $D = 2$, an individual $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ is composed by a sequence of genes (e.g., $\{x_1, y_1, z_1, x_2, y_2, z_2\}$).

7.1.2. Algorithm Steps. The main steps followed by our GA can be summarized as follows:

(1) An *initial population* is generated.
(2) Each individual from the generation is *evaluated* and ranked by assessing a *raw fitness score* (e.g., objective function of the problem (3)).
(3) Some individuals are *selected* to be parents according to their position in the ranking.
(4) A new generation is created as follows:
  (i) A 5% of the new individuals are a copy of the top 5% of the previous generation (elite individuals).
  (ii) Of the remaining individuals:
    (a) 80% is created by *Crossover-and-Mutation* (CM) combining the genes of two selected parents (crossover) and applying a mutation to these new individuals with a very low probability p_m^{CM}.
    (b) 20% is created by an operation of mutation of some parents (termed self-reproduction and mutation, SRM) with probability p_m^{SRM}.
(5) The *exit criteria* are checked after creating the new generation. The algorithm finishes when the lowest raw fitness score found cannot be improved (i.e., lowered) after 50 consecutive generations by at least 10^{-3}. If this condition is not met, go to step (2). However, if within the first MAX_G generations no individual meets problem constraints, the algorithm returns $\mathcal{L} = \varnothing$ to indicate that no solutions can be found with that number of drones.

For example, for a population size of 200 individuals, each generation would be composed of 10 elite individuals (5%), 152 individuals formed by CM, and 38 individuals formed by SRM. Although this differs slightly from the traditional sequential application of operations done by the canonical GA, it improves convergence according to [48, 49].

7.1.3. Initial Population. An initial population of p individuals is created. Experimentally, we found that a population size of $p = 200$ individuals provides results that cannot be improved in the scenarios tested.

According to [50], choosing an initial population which is not completely random, but containing individuals prone to be fit, improves the performance of the GA. For this reason, we use the *k-means* clusterization method illustrated in Figure 6. In particular, inside an area similar to the grid X-Y dimensions, U points—which represent 2D user locations from the set $\{\boldsymbol{w}_k \mid k \in \mathcal{U}\}$—are taken as input to create D clusters so that the mean distance from each point to its cluster centroid is minimized. Once the *k-means* method is applied, we add to each 2D centroid the average altitude (i.e., vertical axis) of the grid, obtaining a set of 3D centroids. The first individual of the initial population will locate its drones in the grid edges closest to these centroids.

The other $(p - 1)$ individuals of the initial population are created by distributing each individual's drones randomly among the spatial regions created around each centroid. In particular, we define cubes with a volume equal to the d-th part of the grid volume. Observe that since the centroid of each cube was determined by k-means clusterization, there could be overlapping regions or regions out of the grid bounds such as those illustrated (in 2D) in Figure 6. Note that drones are always located at grid edges.

To guarantee convergence to the optimal, after creating the initial population, these spatial regions are no longer taken into consideration for next generations. This implies that genes carried by individuals from next generations could be any edge $\in \mathcal{P}$ without any restriction.

7.1.4. Operators. The following operators have been used:

(1) *Evaluation* and ranking: each individual is assigned a raw fitness score by assessing the problem objective function as detailled in Appendix C. Since the genetic algorithm is called to search for solutions with D

Table 4: Heuristics validation input parameters.

IEEE Standard		Scenario		Traffic		Constraints	
Revision	802.11n	Users	$\{10, \ldots, 40\}$	Call length	180 s	$RSSI_{min}$	-82 dBm
GI	800 ns	Size	50 m × 50 m	VoIP codec	G.711	SNR_{min}	20 dB
Preamble	Greenfield	X-Y step	1 m	On/Off times	CBR	R_{min}	65
Bandwidth	20 MHz	Altitude layers	20 m	Packet interval	20 ms		
Retries	7	Prop. Exponent	3.3	Broadcast channels	1		

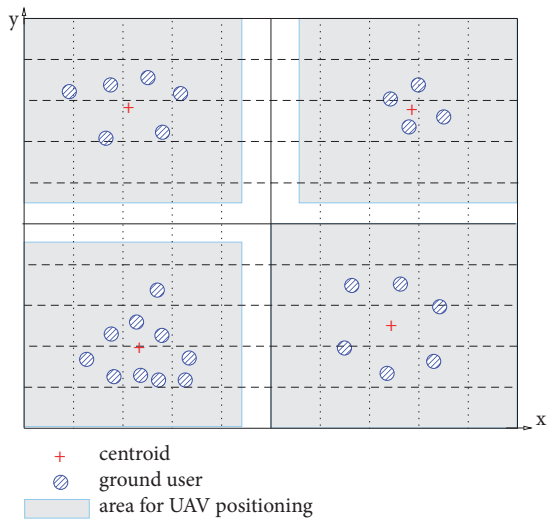


Figure 6: Example of 2D spatial regions where drones of initial population are confined ($D = 4$).

drones, individuals that do not meet the problem constraints are assigned a higher score by adding a penalty (a number between 1 and 2, according to the ratio of ground users that meet QoS constraints) to the value of their objective function. Individuals are then sorted according to their raw fitness score. Finally, if an individual is in the n position in the ranking, it is assigned a new scored termed expectation value of $1/\sqrt{n}$.

(2) Parents *Selection:* we use a stochastic uniform selection process among individuals according to their expectation value. In our case, 342 parents (152 × 2 CM + 38 SRM) are selected for crossover and/or mutation operations. Therefore, individuals in the top positions are chosen multiple times to be parents.

(3) *Crossover:* the crossover operator combines the characteristics of two parents to create a new individual. We use a uniform crossover operation by generating a random binary vector which determines for each gene of the child whether it comes from one parent or the other.

(4) *Mutation:* We apply an exchange-type uniform mutation that consists of changing one gene of an individual with a given probability (p_m^{CM} for those individuals generated after crossover, or p_m^{SRM} for parents) for a random coordinate within the grid edges (we set $p_m^{SRM} = 1/(3 \cdot D)$ and $p_m^{CM} = 1/2 \cdot p_m^{SRM}$).

7.2. Heuristics Performance. This section validates the heuristic method proposed above in terms of accuracy and convergence speed. To do so, we propose a series of experiments using the parameters listed in Table 4. The number of users has been increased from 10 to 40 in a fixed area of 50 m × 50 m and as a result, the number of required drones has also increased. Each experiment has been repeated 30 times (users' position is generated randomly on each run). Results shown represent averaged values.

7.2.1. Quality of the Solutions. The optimal solution found with our heuristic method is compared with that from exhaustive search and from simply applying the k-means clusterization method (i.e., the first individual in our initial population). Table 5 shows for each method the number of drones D, the average energy consumption at the stations as E, and the value of the problem objective function. Results also include the standard deviation between parentheses when greater than 10^{-2}. It can be observed that, while results obtained with our heuristic method are very close to the optimal ones (i.e., provided by exhaustive search), the k-means clusterization method always overestimates the number of required drones.

Results have been extended from 50 to 150 ground users for our heuristic method and k-means but not for the exhaustive search one (such extension poses a computational burden too high for exhaustive search method with the grid under consideration. For example, for $D = 3$ and $|\mathscr{P}| = 2\,500$ edges, more than $15 \cdot 10^9$ possible UAV location combinations would have to be evaluated). Figure 7 plots the value of the objective function in our extended comparison. Results suggest that the number of drones obtained with our heuristics tends to grow linearly with the density of users in the scenario under consideration, outperforming the solutions found with k-means in the studied cases. Observe that since the energy term in (3) is normalized and typically small, the objective function value is almost entirely determined by the number of drones (which explains the steps in the plot).

7.2.2. Convergence Speed and Complexity. The number of generations until our heuristic algorithm finds the solution is plotted in Figure 8. Notice that if no individual satisfies the problem constraints during MAX_G consecutive generations, the number of drones D is increased and the *GAsearch* procedure is called again (this value was determined experimentally: initially, we tried $MAX_G = 200$ generations in an scenario of 10 000 m with 1 m X-Y steps but, since solutions were always found during the first five generations,

Table 5: Comparison of exhaustive, heuristic, and k-means.

U		10	20	30	40
exhaustive	D	1	1	2	2
	E	0.24	0.34	0.28	0.33
	of	1.10	1.14	2.12	2.13
genetic	D	1	1	2	2
	E	0.24	0.34	0.29	0.33
	of	1.10	1.14	2.12	2.13
k-means	D	3 (0.62)	4 (0.88)	5 (0.96)	5 (1.07)
	E	0.17 (0.01)	0.20 (0.02)	0.21 (0.02)	0.22 (0.02)
	of	3.47 (0.62)	4.38 (0.87)	5.18 (0.95)	5.49 (1.06)

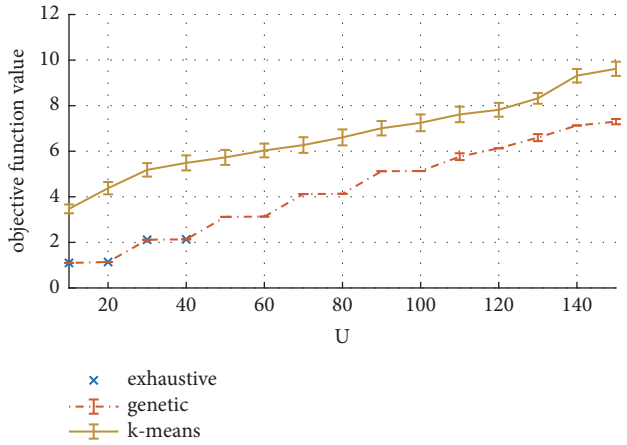


Figure 7: Comparing heuristics, exhaustive search, and clusterization.

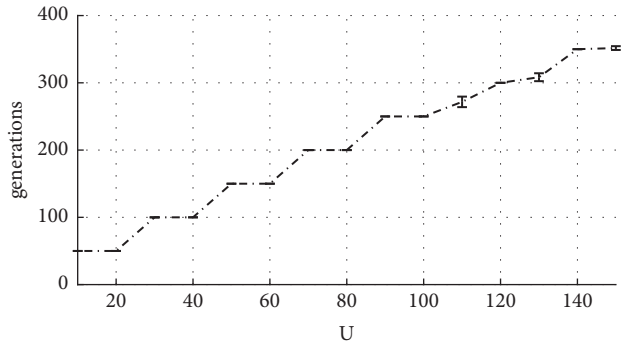


Figure 8: Generations evaluated until exit criteria are met.

we decided to reduce it to 50.). Results show that the best individual is always found within the first five generations in all scenarios tested. This explains that solutions with D drones (see Figures 7 and 8) commonly require evaluating around $\sim 50D$ generations (see our exit criteria).

The fast convergence speed obtained can be explained as follows:

(i) Both the initial population selection based on applying k-means and the large population size ($p = 200$) create a first generation of individuals prone to be excellent candidates.

(ii) The number of grid edges ($|\mathscr{P}|$) (which depends on the discretization step and the terrain size) also impacts on convergence speed. Observe that the number of grid edges that belong to each cluster (i.e., $|\mathscr{P}|/D$) is not too large when compared to the population size p. In general, scenarios with a greater number of grid edges would need a larger population to attain fast convergence.

In terms of computational complexity, the most demanding procedure in our heuristic method is the assessment of the raw fitness score (see Appendix C), whose computation time can be expressed as $O(U \cdot D)$. However, if one considers that D is upper bounded by the constant $D_{\max}$ and that the number of evaluations of the fitness function is also upper bounded in our heuristic, then the computational complexity of our heuristic method would be $O(U)$ (if the upper limit of D was not bounded by the problem constraint but by the number of ground users U, then the computational complexity would be $O(U^3)$. At any rate, solutions can be computed in polynomial time.). Figure 9 shows the execution time obtained in the previous experiments with a computer with 16 GB RAM and CPU Intel Broadwell x86 (8 cores and 4 threads per core) at 2.2 GHz. Results confirm computation time monotonically increases with U in less than polynomial time for our heuristic method. The exhaustive search method, however, shows a stepped rise in computation time when the number of drones in the solution is increased from 1 to 2 at $U = 30$ (see Figure 7).

From the previous results it can be concluded that the quality of the solutions found with the proposed heuristic is acceptable and that computational complexity and computation time are also acceptable for practical purposes.

8. Numerical Results

In this section we provide a numerical analysis using the heuristics proposed in Section 7. All results shown represent the average value after repeating each scenario 30 times (each time the user disposition was random). We set the focus on three different aspects: the impact of the IEEE 802.11 standard in use in the number of UAVs used, the energy

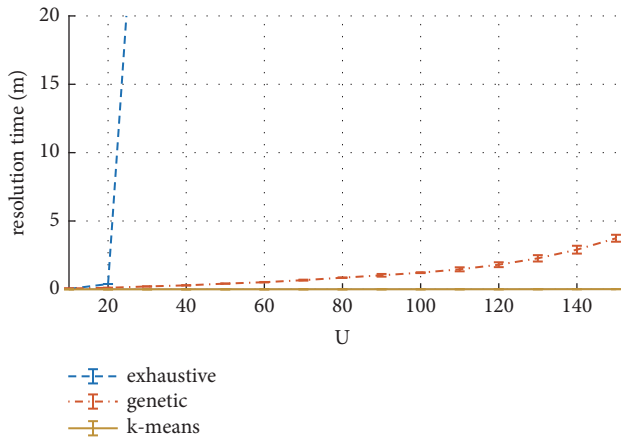


Figure 9: CPU execution time obtained for the methods used.

Table 6: IEEE 802.11 revisions considered.

Frequency	Channel Bandwidth	Revision
2.4 GHz	20 MHz	802.11g
	20 MHz	802.11n
	40 MHz	
5 GHz	20 MHz	802.11ac
	80 MHz	

consumption and its dependence of traffic load, and the impact of speech quality and grid bounds constraints in the energy consumption of ground users.

8.1. Influence of the IEEE 802.11 Standard Revision. In this section, we compare the performance of the widely adopted IEEE 802.11 g/n/ac standard amendments with the configuration listed in Table 6. In particular we would like to find out if certain amendments are more fit than others for the two following scenarios of application:

(i) *Low user density* scenario: for a fixed number of users ($U = 100$) we change the terrain size from 100 to 10 000 m^2, obtaining a range of user sparsity from 1 to 100 m^2/user. In this type of scenarios the number of UAVs required is expected to increase mainly due to the signal coverage requirement.

(ii) *High user density* scenario, (sparsity $\leq 1 m^2$/user): in a small area of 25 m^2 we change the number of users from 25 to 100 which yields a user sparsity range from 1 to 0.25 m^2/user. In these circumstances the number of drones required is mainly determined by the need to satisfy the Speech Quality constraint (R_{min}).

Results are shown in Figures 10(a) and 10(b). In the low user density, large area scenario (Figure 10(a)), it can be observed that those standard revisions using the 2.4GHz band (g and n) outperform (for large areas) the revision ac which uses 5GHz as a result of lower path loss. Regarding the small scenario with high-density shown in Figure 10(b), no significant differences can be observed among different revisions.

8.2. NIC Energy Consumption. We study how the area size and user density impact on the average energy consumption of ground users. In our study we define two terrain sizes of 10 m × 10 m and 100 m × 100 m (100 m^2 and 10 000 m^2 respectively), changing the number of ground users from 10 to 100 in steps of 10. The energy is expected to increase with the number of users associated per drone (i.e., $|\mathscr{C}(k)|$) as a result of an increment in collisions.

Results shown in Figure 11 represent the average energy consumption of ground users and the UAV altitude from all drones. The following points can be made after examining these results:

(i) In the small size scenario, at some points the number of users increase but the average energy decreases (e.g., between 20 and 30 users in Figure 11(b)) which seems anti-intuitive. This is attributable to an increment in the number of UAVs (see Figure 11(a)), which is consistent with our hypothesis since the number of users associated with each AP decreases due to the increment in the number of UAVs.

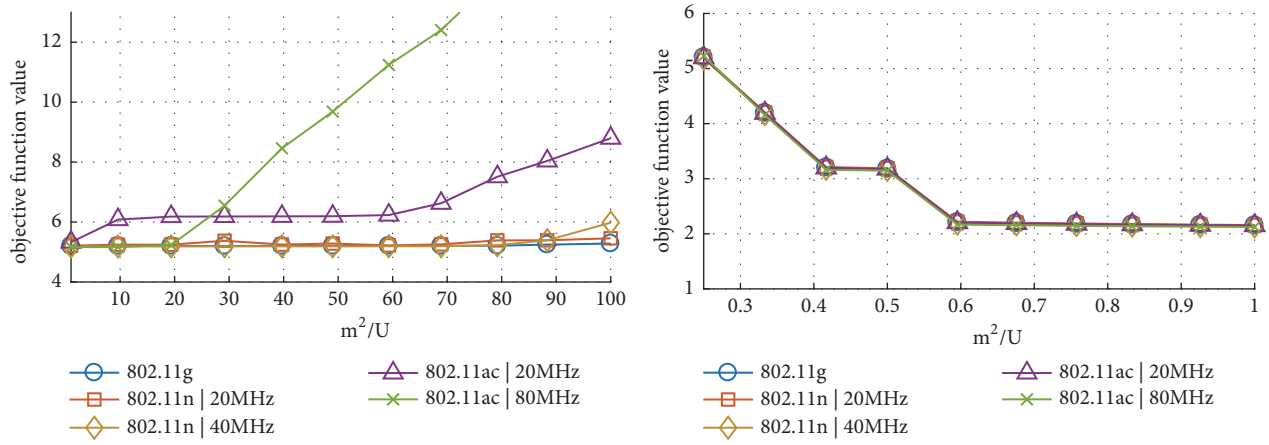
(ii) Ground users in the smaller size terrain exhibit higher energy expenditure than those in the bigger size up to a number of 60 users. This can also be explained as a result of less drones being deployed and a higher number of users associated with each drone, and therefore more traffic load per drone is generated.

(iii) Both curves get slightly inverted after 70 users. In this case the number of deployed UAVs is the same in both scenarios, so energy consumption in both cases tend to be more similar although a little higher in the bigger terrain. This can be explained looking at the average drone altitude in Figure 11(c). The larger area produces UAVs in higher altitude to provide coverage. As a consequence, path loss is higher and more users receive a weaker signal, setting a MCS that sets a lower data bit-rate, which translates into more transmission time.

8.3. Relation between Energy and Constraints: R_{min} and Grid Bounds. We would like to show how the minimum speech quality (R_{min}) and the maximum height of the grid (h_{max}) used in the problem have a direct impact in the energy consumption of ground users. We perform a different experiment for each constraint:

(i) Limited altitude scenario: we create a large area (10 000 m^2) with a range ground users from 10 to 100 and limit the maximum altitude of the grid to ($h_{max} = \{10, 25, 40\}$ meters). In this scenario it is expected that the low upper bound results into more UAVs in the solutions and, in turn, less users per drone which should translate into less energy consumption.

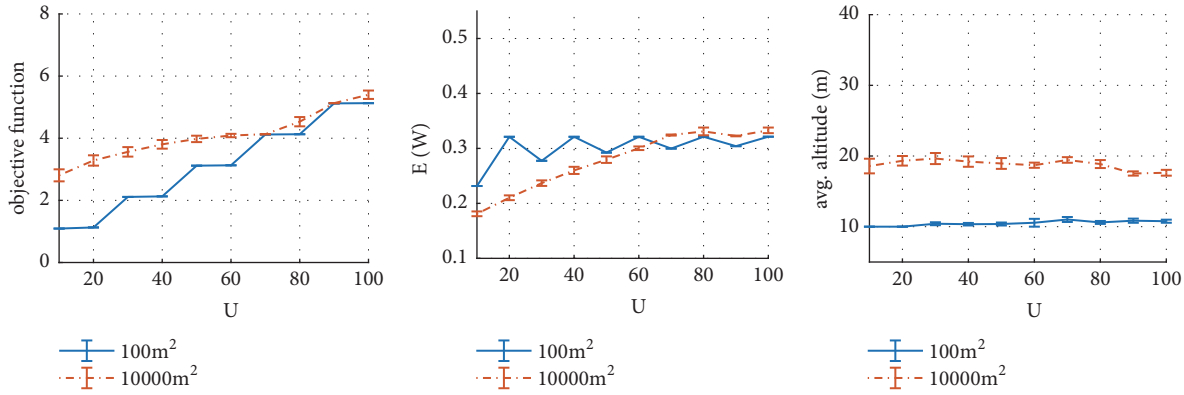
(ii) Higher speech quality scenario: we create a small area of 100 m^2 with a range of ground users from 10 to 100 so that coverage is not demanding. Then we



(a) Low user density, large area scenario

(b) High user density, small area scenario

Figure 10: Comparison of various revisions of standard IEEE 802.11.



(a) Objective function

(b) Average energy consumption

(c) Average UAV altitude

Figure 11: Energy consumption analysis.

solve our deployment problem for different values of the speech quality constraint ($R_{\min} = \{70, 75, 80\}$). It is expected that higher quality turns into more UAVs and consequently, as previously stated, should translate into less energy consumption.

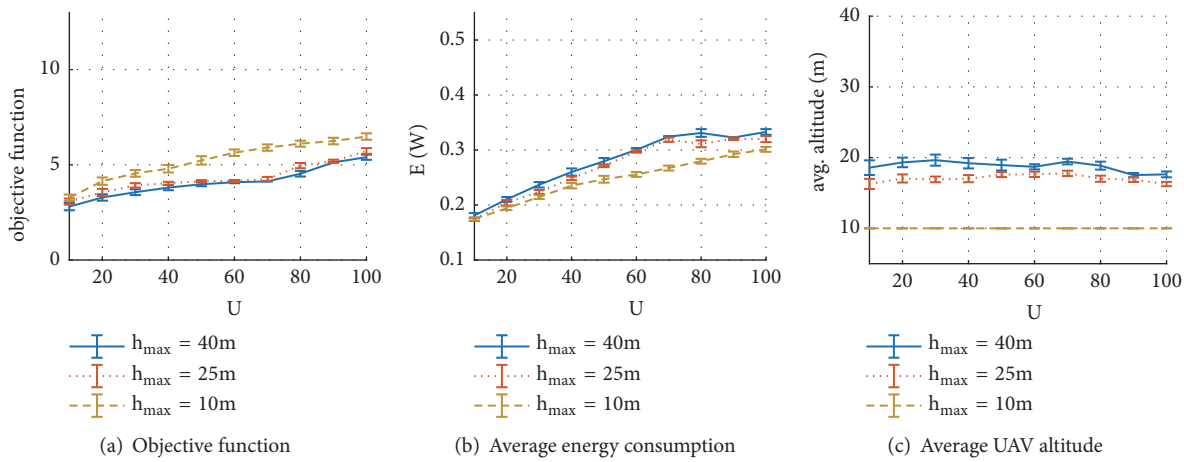
The results obtained in both scenarios are shown in Figures 12 and 13 respectively. Results suggest that our rationale is confirmed: the greater maximum height is, the less UAVs are launched and as a consequence the energy expenditure in ground user stations tends to decrease. The same logic applies to decreasing the minimum speech quality $R_{\min}$. Choosing one or the other constraint to regulate the energy consumption has various implications. Limiting the altitude in the grid bound will be more effective in large areas where users are disperse, while increasing the minimum speech quality will have more impact in scenarios with higher density of ground users. However, note than choosing very strict bounds (i.e., too high quality or too low altitude) can severely overestimate the number of UAVs to be deployed.

9. Discussion of the applicability to SAR Missions and Open Issues

As stated in Section 1, this work is a first step into the provision of a WiFi network that enables a reliable communication service for SAR missions. The problem defined in Equation (3) is aimed at solving the initial deployment of UAVs, but it can also be applied to provide service to ground users gathered in specific areas such as meeting points or rescue areas, or in situations when ground users remain semistatic (e.g., users are conducted to remain still by the communication broadcast channel). In this section we provide a brief analysis of its applicability when ground users are moving and identify key issues to be addressed in further research.

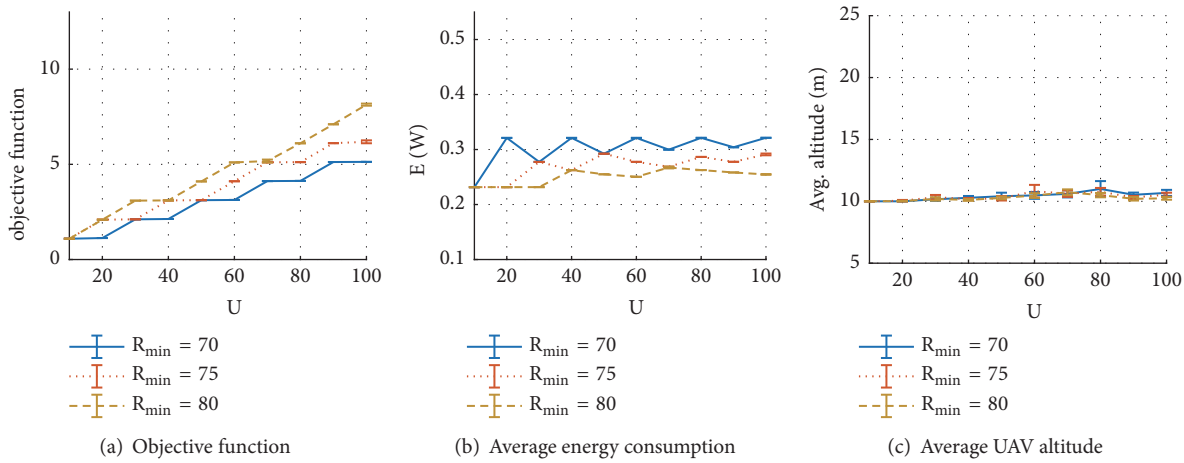
9.1. Conditions for the Validity of the Solutions. Given a solution obtained at time t_1, the conditions for its validity in t_2 (being $t_1 < t_2$) are

(i) $C_i(t_1) = C_i(t_2)$, $\forall i \in \mathcal{D}$ (i.e., wireless clients do not change their AP)



(a) Objective function (b) Average energy consumption (c) Average UAV altitude

Figure 12: Experiment: limited altitude scenario.



(a) Objective function (b) Average energy consumption (c) Average UAV altitude

Figure 13: Experiment: higher speech quality scenario.

(ii) $RSSI_{ij}(t_1) \leq RSSI_{ij}(t_2)$, $\forall i, \forall j \in \mathcal{U}$ (i.e., ground station's data rate, energy, and speech quality are not negatively affected)

When ground users can freely move the previous conditions are unlikely to be satisfied after a while. In such case, a new solution search has to be carried out, arising new issues to be discussed next.

9.2. Preliminary Analysis of the Applicability When Users Are Moving. Assuming that ground users can be tracked, their position could be periodically checked in a process such as the one illustrated in Figure 14 where T represents the period between consecutive observations (an alternative scheme to Figure 14 is to ignore validity conditions and recalculate new solutions every T seconds.). In such process, the following key points have to be considered:

(i) The observation period (T): it should be set according to how ground users are expected to move (e.g., speed and direction). Note that the difference between the theoretical set of user locations $\{\boldsymbol{w}_k\}$ at the beginning of each observation period and the actual position of ground users at the end of such period will likely grow with T. As such, long values of T will produce less accurate speech quality prediction and higher chances of not satisfying the validity conditions. Although in Figure 14 we assume that T is long enough to relocate UAVs, its strict lower bound simply requires $T \geq T_{\text{sol}}$, where T_{sol} stands for the time required to compute a new solution. Obviously, the shorter computation delay the better, which in turn depends on the computational resources available as well as the heuristic method used to find solutions (our heuristic method can be sped up in various ways and other alternative heuristics could also be explored).

(ii) Relocating UAVs: relocating drones may result in transient periods where the problem constraints are not met for some users. The actual strategy for UAVs displacement constitutes a problem itself that deserves further analysis as it should minimize not only the relocation delay (e.g., see [51]) but also service disruption to ground users. For example, initial

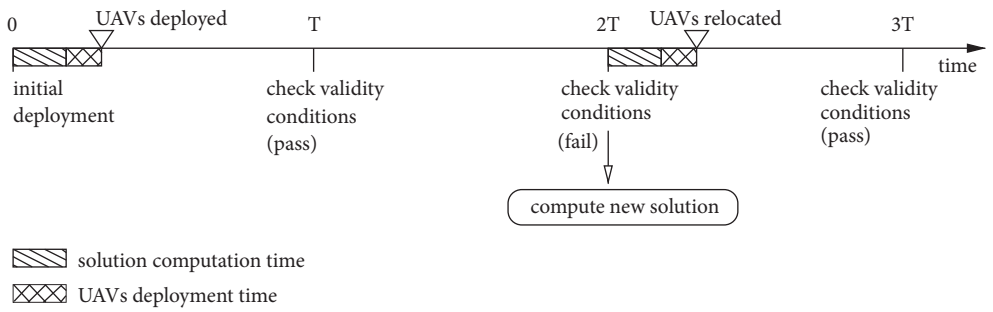


Figure 14: Continuous process for the provision of the communication service.

strategies could include minimizing the sum of all UAVs displacements, the longest UAV displacement, or the number of users affected by service outages.

(iii) Problem objective function: after the previous discussion it seems reasonable to define a new problem targeted to the relocation of UAVs during service provision. This new problem could be an extension of the one defined in this paper but including a cost that accounts periods of disrupted communication, which allows one to balance the pros and cons of relocating UAVs. For example, it may be reasonable to keep the last solution as long as the problem constraints are satisfied if the new solution implies a transient period where some users lose their connectivity.

For illustrative purposes and as a prospective exercise, we have performed the following experiment: using the scenario and parameters shown in Table 4 we have placed users who move according to a correlated random walk model (each ground user moves at a walking speed of 5.3 km/h with a probability of 0.8 and, in such case, the direction remains unchanged with probability 0.8. Users rotate 180 degrees when they reach the area bounds.). Then, we perform an off-line computation (i.e., $T_{sol} = 0$) of the optimal solution every T seconds. Algorithm 2 is used for the initial deployment and the number of drones deployed D remains unchanged for the remainder of the simulation (30 min). When recalculating new solutions we simply get the best solution possible with D drones found by our GA (even though problem constraints were not satisfied). With every new solution, drones are relocated (considering a speed of 60 Km/h) so that the sum of the distances travelled by all drones is the minimum possible. Every five seconds the position of ground users is sampled and the problem constraints (i.e., coverage and speech quality) are verified. For each user, we define a service disruption event when the speech quality constraint is not satisfied for two consecutive samples.

Figure 15(a) shows the percentage of the simulation time (i.e., 30 minutes) that a ground user on average experiences a service free of disruption events. As expected, results show that longer observation periods provide worse performance, obtaining uneventful service during more than 90% of the time for $T \leq 60$ seconds. In our simulations, the longest distance ever travelled by any drone was always shorter than 20 meters, which results in a maximum relocation delay of 1.2 seconds.

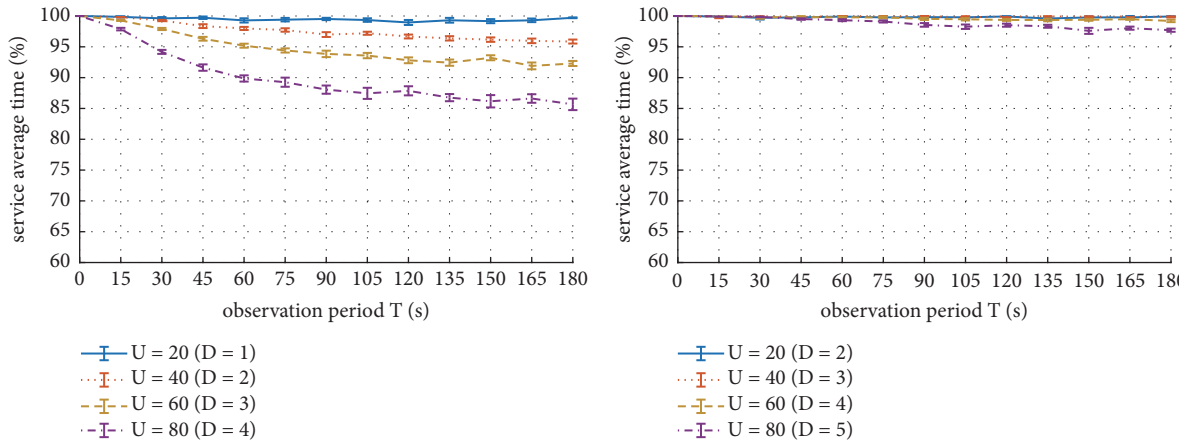
A final experiment was done similar to the previous one but launching one extra drone ($D + 1$) in the initial deployment (i.e., one extra iteration in the loop in Algorithm 2). The results obtained are plotted in Figure 15(b) and suggest that launching more drones than strictly necessary ($D + 1$ in our case) can be an effective method to avoid service degradation during transient periods.

10. Conclusions and Further Work

In this paper we have addressed the problem of UAVs deployment to provide a reliable emergency communication service usable in SAR missions. We have formulated a new problem suited for the initial deployment of UAVs which can also be applied to provide service to ground users gathered in specific areas such as meeting points or rescue areas or in situations when ground users remain semistatic. We have also analyzed the limits of its applicability to moving users. Our optimization problem minimizes the number of drones required to provide a reliable communication service. Among equivalent solutions we choose the one that minimizes energy expenditure at ground user's stations due to communication. Optimal solutions are found using well-known metaheuristic such as GA since exhaustive search is not generally feasible due to its high computational complexity.

Numerical results suggest large area scenarios with highly disperse users benefit from IEEE standards using 2.4 GHz and the lowest bandwidth possible, as less drones will be needed to provide the service. The average number of ground users associated with each UAV will also have impact in the energy consumption and, therefore, in the ground users battery life. For that reason, problem constraints such as the maximum altitude allowed to UAVs or the minimum speech quality required can be used to increase energy expenditure at ground user stations.

A number of open issues linked to users' mobility have been pointed for further research. In addition to including the cost of relocating drones, other optimization problems can also be studied in the context of SAR missions (e.g., optimizing drones' energy expenditure, maximizing the number of covered users given a fixed number of drones, etc.). Relocation strategies should also be explored, considering not only the minimization of the deployment delay but also the



(a) Launching D drones using the optimal solution

(b) Launching optimal solution plus one extra drone

Figure 15: Experiment: percentage of time with uneventful communication per average user.

number of users affected and the level of degradation of the service.

Appendix

A. Expected Time per Slot: $E[T]$

The expected time per slot represents the expected length of each state of the Markov chain modelling the IEEE 802.11 system. To come up with a closed-form expression of $E[T]$, we follow the approach presented in [34], which considers the probability of being on each of the following states:

(i) Idle state, when nobody attempts to transmit.

(ii) Success state, when only a single station attempts to transmit with no channel related errors.

(iii) Error state, when a single station attempts to transmit but channel errors occur.

(iv) Collision state, when more than one station attempts to transmit simultaneously (same slot).

Consequently, we can obtain $E[T]$ by weighting the time spent on each state as shown in (A.1).

$$E[T] = T_I + T_S + T_E + T_C \tag{A.1}$$

In the previous equation, T_I can be calculated as follows:

$$T_I = P_i \sigma \tag{A.2}$$

where σ stands for the timeslot duration as defined in the corresponding IEEE 802.11 revision. P_i is the probability that the channel is not busy in a randomly chosen slot, which can be expressed as

$$P_i = \prod_{j=1}^{S} \left(1 - \tau^{(j)}\right) \tag{A.3}$$

where S was the number of stations in the system (including the AP).

The time spent on successful (T_S) and erroneous (T_E) slots in (A.1) are as follows:

$$T_S = \sum_{j=1}^{S} P_s^{(j)} \left(1 - \mathrm{FER}^{(j)}\right) T_s^{(j)} \tag{A.4}$$

$$T_E = \sum_{j=1}^{S} P_s^{(j)} \mathrm{FER}^{(j)} T_e^{(j)} \tag{A.5}$$

where $T_s^{(j)}$ and $T_e^{(j)}$ represent the average time that station j spends in successful and erroneous transmissions, respectively. Their calculus depend on the used standard and physical data rate due to the fact that every waiting interframe periods (DIFS, SIFS...) and time spent in transmission (headers, payload, ack.) must be taken into account (note that in the case of the access point, these times may be averaged considering every conversation with different data rates and/or codecs). In the previous equations, $P_s^{(j)}$ stands for the probability that only an observed station j attempts to transmit while the rest remain silent, which can be expressed as

$$P_s^{(j)} = \tau^{(j)} \prod_{k=1, k\neq j}^{S} \left(1 - \tau^{(k)}\right) \tag{A.6}$$

Finally, the time spent in collided transmissions is represented by T_C in (A.1). Each station has an average time for collision, $T_c^{(j)}$, according to the time spent in an erroneous transmission (i.e., $T_c^{(j)} = T_e^{(j)}$). Nevertheless, when packets sent by two different stations suffer collision, the time to be considered corresponds with the longest average collision time. In order to address this concern we will group stations in traffic classes according to its $T_c^{(j)}$ (all stations belonging to the same class have the same average collision time). We consider N_c traffic classes tagged with $d \in \{1, \ldots, N_c\}$ ordered from higher to lower channel occupancy during collisions. Hence, we define $C(d)$ as the set of stations that belongs to class d.

Now we can calculate the probability that at least one station from the *d-th* class transmits

$$P_{tx}^{C(d)} = 1 - \prod_{j \in C(d)} \left(1 - \tau^{(j)}\right) \tag{A.7}$$

as well as the probability that at least one station from a higher or lower class transmits

$$P_{tx}^{H(d)} = 1 - \prod_{i=d+1}^{N_c} \prod_{j \in C(i)} \left(1 - \tau^{(j)}\right) \tag{A.8}$$

$$P_{tx}^{L(d)} = 1 - \prod_{i=1}^{d-1} \prod_{j \in C(i)} \left(1 - \tau^{(j)}\right) \tag{A.9}$$

Due to the fact that lower classes slow down higher ones as a result of longer transmissions, T_C can be calculated as

$$T_C = \sum_{d=1}^{N_c} \left(P_c^{C(d)} + P_c^{H(d)}\right) T_c^{(d)} \tag{A.10}$$

where $T_c^{(d)}$ is the average time that any station from class d spends on a collided transmission and $P_c^{C(d)}$ represents the probability that any collision occurs between stations from the same class d:

$$P_c^{C(d)} = \left(1 - P_{tx}^{H(d)}\right) \cdot \left(1 - P_{tx}^{L(d)}\right) \cdot \left(P_{tx}^{C(d)} - \sum_{j \in C(d)} P_s^{(j)}\right), \tag{A.11}$$

and $P_c^{H(d)}$ accounts for the probability that class d is involved in a collision with at least one station from a higher class.

$$P_c^{H(d)} = P_{tx}^{C(d)} \cdot P_{tx}^{H(d)} \cdot \left(1 - P_{tx}^{L(d)}\right) \tag{A.12}$$

A closed-form expression of $E[T]$ is obtained with the proper substitution of T_I, T_S, T_E, and T_C in (A.1).

B. Expected Energy Consumption per Slot: $E[J]$

During a slot, different events can happen with a certain probability each one with different energy consumption. Then it is possible to estimate the average energy consumption during a slot as specified in expression (24). This appendix details how to calculate each term of such equation.

B.1. Probabilities Associated with Each Event. The channel is idle with a probability of

$$P_i = \prod_{j=1}^{S} \left(1 - \tau^{(j)}\right) \tag{B.1}$$

Probability of successful packet transmission by station j:

$$P_{tx,s}^{(j)} = \left(1 - \mathrm{FER}^{(j)}\right) \cdot P_s^{(j)} \tag{B.2}$$

where $P_s^{(j)}$ is defined in (A.6). Probability of transmitting a corrupted packet by station j:

$$P_{tx,e}^{(j)} = \mathrm{FER}^{(j)} \cdot P_s^{(j)} \tag{B.3}$$

Probability of collision by station j

$$P_{tx,c}^{(j)} = \tau^{(j)} \cdot \left(1 - \prod_{i=1, i \neq j}^{S} \left(1 - \tau^{(i)}\right)\right) \tag{B.4}$$

Probability of receiving a successful packet (destinated to station j) by station $j \neq ap$:

$$P_{rx,s}^{(j)} = P_{tx,s}^{(ap)} \frac{\lambda_{rx}^{(j)}}{\lambda^{(ap)}} \tag{B.5}$$

Probability of listening to a successful packet (with destination other station) by station j:

$$P_{o,s}^{(j)} = \left(\sum_{i=1, i \neq j}^{S} P_{tx,s}^{(i)}\right) - P_{rx,s}^{(j)} \tag{B.6}$$

Probability of listening to a corrupted packet (to any destination) by station j:

$$P_{rx,e}^{(j)} = \sum_{i=1, i \neq j}^{S} P_{tx,e}^{(i)} \tag{B.7}$$

Probability of listening to a collision (any destination) by station j:

$$P_{rx,c}^{(j)} = 1 - P_i - P_{tx,c}^{(j)} - \sum_{i=1}^{S} P_s^{(i)} \tag{B.8}$$

B.2. Energy Consumption Terms Associated with Each of the Previous Events

$$J_\sigma = \rho_{idle} \sigma \tag{B.9a}$$

$$J_{tx,s}^{(j)} = \rho_{tx} T_{tx,data}^{(j)} + \rho_{rx} T_{ack} + \rho_{idle} (\mathrm{SIFS} + 2\delta + \mathrm{DIFS}) \tag{B.9b}$$

$$J_{tx,c}^{(j)} = \rho_{tx} \overline{T}_{data|c} + \rho_{idle} (\mathrm{EIFS} + \delta) \tag{B.9c}$$

$$J_{tx,e}^{(j)} = \rho_{tx} \overline{T}_{data|e} + \rho_{idle} (\mathrm{EIFS} + \delta) \tag{B.9d}$$

$$J_{rx,s}^{(j)} = \rho_{rx} T_{rx,data}^{(j)} + \rho_{tx} T_{ack} + \rho_{idle} (\mathrm{SIFS} + 2\delta + \mathrm{DIFS}) \tag{B.9e}$$

$$J_{rx,c}^{(j)} = \rho_{rx} \overline{T}_{data|c} + \rho_{idle} (\mathrm{EIFS} + \delta) \tag{B.9f}$$

$$J_{rx,e}^{(j)} = \rho_{rx} \overline{T}_{data|e} + \rho_{idle} (\mathrm{EIFS} + \delta) \tag{B.9g}$$

$$J_{o,s}^{(j)} = \rho_{rx} \overline{T}_{data|s} + \rho_{rx} T_{ack} + \rho_{idle} (\mathrm{SIFS} + 2\delta + \mathrm{DIFS}) \tag{B.9h}$$

where

$$T_{tx,data}^{(j)} = \text{PLCP} + \frac{\left(L_{header}^{(j)} + L_{data}^{(j)}\right)}{R_b^{(j)}} \tag{B.10a}$$

$$T_{rx,data}^{(j)} = \frac{T_{tx,data}^{(j)} + B_{CH} \cdot T_{tx,data}^{(broadcast)}}{1 + B_{CH}} \tag{B.10b}$$

$$T_{ack} = \text{PLCP} + \frac{L_{ack}}{R} \tag{B.10c}$$

where R_b (physical data bit-rate) depends on the MCS, R corresponds to the lowest bit-rate (MCS 0), B_{CH} represents the number of broadcast channels, and $T_{tx,data}^{(broadcast)}$ represents the transmission time of data.

In the previous equations ρ_{tx}, ρ_{rx}, and ρ_{idle} represent the power consumption on each power state in the NIC; $T_{data}^{(j)}$ and $T_{ack}^{(j)}$ indicate the duration of the reception/transmission of a data packet or ACK at station j; $\overline{T}_{data|c}$, $\overline{T}_{data|e}$, and $\overline{T}_{data|s}$ represent the average duration of the transmission of a data packet that has collided, suffered error, or successfully transmitted, respectively.

$$\overline{T}_{data|s} = \frac{1}{\sum_{j=1}^{S} P_{tx,s}^{(j)}} \cdot \sum_{j=1}^{S} P_{tx,s}^{(j)} T_{tx,data}^{(j)} \tag{B.11a}$$

$$\overline{T}_{data|e} = \frac{1}{\sum_{j=1}^{S} P_{tx,e}^{(j)}} \cdot \sum_{j=1}^{S} P_{tx,e}^{(j)} T_{tx,data}^{(j)} \tag{B.11b}$$

The expression of $\overline{T}_{data|c}$ is more complex as if a collision between two stations occurs, it must preserve the longest duration collision time (the slowest stations negatively affect the fastest ones). To capture this effect we take the terminology from Appendix A, where $d \in \{1, \ldots, N_c\}$ represents different traffic classes in the set of stations considered.

$$\overline{T}_{data|c} = \frac{\sum_{d=1}^{N_c} \left(P_c^{C(d)} + P_c^{H(d)}\right) T_{tx,data}^{(d)}}{\sum_{d=1}^{N_c} \left(P_c^{C(d)} + P_c^{H(d)}\right)} \tag{B.11c}$$

C. Raw Fitness Criteria Evaluation

Algorithm 3 detaills how the raw fitness score has been evaluated. In sum, It requires solving the following tasks:

(i) First, signal constraints are calculated by computing every RSSI and SNR for each possible user-drone association (lines (1)-(6)). Then, each user station is associated with its best drone (lines (7)-(15)).

(ii) The set of stations is created for each drone (i.e., each drone mounts an AP and consequently forms its own network) (line (18)).

(iii) The performance for each IEEE 802.11 network is evaluated by solving the nonlinear equation system iteratively (lines (22)-(36)). The output is the quality factor R and the average energy consumption $\overline{E}$.

Input: $\mathcal{U}$, $\mathcal{D}$, $\{\boldsymbol{w}_i\}$, $\{\boldsymbol{x}_k\}$
Initialization: $f_{\text{fitness}} = D + 2$
Output: f_{fitness}
(1) **for** $i = 1$ **to** U **do**
(2) **for** $j = 1$ **to** D **do**
(3) *solve* $\text{RSSI}_{i,j}$; // Eq. (4)
(4) *solve* $\text{SNR}_{i,j}$; // Eq. (6)
(5) **end**
(6) **end**
(7) **for** $i = 1$ **to** U **do**
(8) **for** $k = 1$ **to** D **do**
(9) **if** $RSSI_{i,k} \geq RSSI_{\min}$ ***and*** $SNR_{i,k} \geq SNR_{\min}$ **then**
(10) **if** $RSSI_{i,k} > RSSI_{i,j}, \forall j \neq k$ **then**
(11) $\mathcal{C}(k) = \mathcal{C}(k) \cup i$;
(12) **end**
(13) **end**
(14) **end**
(15) **end**
(16) **if** $\sum_{i=1}^{D} |\mathcal{C}(i)| = U$ **then**
(17) **for** $k = 1$ **to** D **do**
(18) *build set* $\mathcal{S} = \mathcal{C}(k) \cup \mathcal{AP}$;
(19) **for** $j = 1$ **to** $|\mathcal{S}|$ **do**
(20) *solve* $T_s^{(j)}, T_c^{(j)}, T_e^{(j)}, \text{FER}^{(j)}$;
(21) **end**
(22) $\tau_{\text{next}}^{(j)} = T_c^{(j)}, \forall j$;
(23) **while** $diff > \Delta$ **do**
(24) $\tau^{(j)} = \tau_{\text{next}}^{(j)}$;
(25) *solve* $E[T]$; // Appendix A
(26) **for** $j = 1$ **to** $|\mathcal{S}|$ **do**
(27) *sol.* $E[B^{(j)}]$; // Eq. (13)
(28) *solve* $r^{(j)}, q^{(j)}$; // Eqs. (11), (12)
(29) *solve* $p^{(j)}$; // Eq. (14)
(30) *solve* $\tau_{\text{next}}^{(j)}$; // Eq. (9)
(31) **end**
(32) $diff = \max\left(\left|\tau_{\text{next}}^{(j)} - \tau^{(j)}\right|\right), \forall j$;
(33) **end**
(34) *solve* $\text{R}(k)$; // Eq. (23)
(35) *solve* $\overline{E}(k)$; // Eq. (28)
(36) **end**
(37) **if** $R(k) \geq R_{\min}, \forall k \in \{1, \ldots, D\}$ **then**
(38) $f_{\text{fitness}} = D + \frac{1}{D} \sum_{k=1}^{D} \frac{\overline{E}(k)}{E_{\max}}$;
(39) **else**
(40) $f_{\text{fitness}} = D + 1 + \frac{1}{D} \sum_{k=1}^{D} \left(\frac{\overline{E}(k)}{E_{\max}} \left[\text{R}(k) < \text{R}_{\min}\right] \right)$;
(41) **end**
(42) **end**

Algorithm 3: Fitness function evaluation.

(iv) Finally, *raw fitness scores* are calculated by applying a fitness function which depends on the objective function we introduced in (3) (lines (37)-(41)).

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] M. Erdelj, M. Król, and E. Natalizio, "Wireless sensor networks and multi-UAV systems for natural disaster management," *Computer Networks*, vol. 124, pp. 72–86, 2017.

[2] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the sky: leveraging UAVs for disaster management," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 24–32, 2017.

[3] C. Luo, J. Nightingale, E. Asemota, and C. Grecos, "A UAV-cloud system for disaster sensing applications," in *Proceedings of the 81st Vehicular Technology Conference (VTC Spring '15)*, pp. 1–5, IEEE, 2015.

[4] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.

[5] M. Deruyck, J. Wyckmans, L. Martens, and W. Joseph, "Emergency ad-hoc networks by using drone mounted base stations for a disaster scenario," in *Proceedings of the 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '16)*, pp. 1–7, IEEE, 2016.

[6] G. Tuna, B. Nefzi, and G. Conte, "Unmanned aerial vehicle-aided communications system for disaster recovery," *Journal of Network and Computer Applications*, vol. 41, no. 1, pp. 27–36, 2014.

[7] R. Siegwart, M. Hutter, P. Oettershagen et al., "Legged and flying robots for disaster response," in *Proceedings of the World Engineering Conference and Convention (WECC '15)*, ETH-Zürich, 2015.

[8] T. Andre, K. A. Hummel, A. P. Schoellig et al., "Application-driven design of aerial communication networks," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 129–137, 2014.

[9] I. Jawhar, N. Mohamed, J. Al-Jaroodi, D. P. Agrawal, and S. Zhang, "Communication and networking of UAV-based systems: classification and associated architectures," *Journal of Network and Computer Applications*, vol. 84, pp. 93–108, 2017.

[10] A. G. Fragkiadakis, I. G. Askoxylakis, E. Z. Tragos, and C. V. Verikoukis, "Ubiquitous robust communications for emergency response using multi-operator heterogeneous networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 13, 2011.

[11] A. M. Hayajneh, S. A. R. Zaidi, D. C. McLernon, M. Di Renzo, and M. Ghogho, "Performance analysis of UAV enabled disaster recovery networks: a stochastic geometric framework based on cluster processes," *IEEE Access*, vol. 6, pp. 26215–26230, 2018.

[12] S. Yahyanejad and B. Rinner, "A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 104, pp. 189–202, 2015.

[13] S. Adams, C. Friedland, and M. Levitan, "Unmanned aerial vehicle data acquisition for damage assessment in hurricane events," in *Proceedings of the 8th International Workshop on Remote Sensing for Disaster Management*, vol. 30, Tokyo, Japan, 2010.

[14] T. Fujiwara and T. Watanabe, "An ad hoc networking scheme in hybrid networks for emergency communications," *Ad Hoc Networks*, vol. 3, Part of Special Issue: Data Communication And Topology Control in Ad Hoc Networks, no. 5, pp. 607–620, 2005.

[15] S. Saha, S. Nandi, P. S. Paul, V. K. Shah, A. Roy, and S. K. Das, "Designing delay constrained hybrid ad hoc network infrastructure for post-disaster communication," *Ad Hoc Networks*, vol. 25, pp. 406–429, 2015.

[16] G. Tuna, T. V. Mumcu, and K. Gulez, "Design strategies of unmanned aerial vehicle-aided communication for disaster recovery," in *Proceedings of the 9th International Conference on High Capacity Optical Networks and Enabling Technologies (HONET '12)*, pp. 115–119, IEEE, December 2012.

[17] Y. Bai, W. Duh, Z. Ma, C. Shen, Y. Zhou, and B. Chen, "Emergency communication system by heterogeneous wireless networking," in *Proceedings of the IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS '10)*, pp. 488–492, June 2010.

[18] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3949–3963, 2016.

[19] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.

[20] D. Zorbas, L. Di Puglia Pugliese, T. Razafindralambo, and F. Guerriero, "Optimal drone placement and cost-efficient target coverage," *Journal of Network and Computer Applications*, vol. 75, pp. 16–31, 2016.

[21] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.

[22] S. Morgenthaler, T. Braun, Z. Zhao, T. Staub, and M. Anwander, "UAVNet: a mobile wireless mesh network using unmanned aerial vehicles," in *Proceedings of the IEEE Globecom Workshops (GC Wkshps '12)*, pp. 1603–1608, IEEE, Anaheim, Calif, USA, December 2012.

[23] E. Yanmaz, R. Kuschnig, and C. Bettstetter, "Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility," in *Proceedings of the 32nd IEEE Conference on Computer Communications (IEEE INFOCOM '13)*, pp. 120–124, April 2013.

[24] C.-M. Cheng, P.-H. Hsiao, H. Kung, and D. Vlah, "Performance measurement of 802.11 a wireless links from uav to ground nodes with various antenna orientations," in *Proceedings of the 15th International Conference on Computer Communications and Networks (ICCCN '06)*, pp. 303–308, IEEE, 2006.

[25] S. Aust, R. V. Prasad, and I. G. M. M. Niemegeers, "Outdoor long-range WLANs: a lesson for IEEE 802.11ah," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1761–1775, 2015.

[26] S. Shin and H. Schulzrinne, "Measurement and analysis of the VoIP capacity in IEEE 802.11 WLAN," *IEEE Transactions on Mobile Computing*, vol. 8, no. 9, pp. 1265–1279, 2009.

[27] E. Charfi, L. Chaari, and L. Kamoun, "New adaptive frame aggregation call admission control (AFA-CAC) for high throughput WLANs," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 3, pp. 469–481, 2015.

[28] H.-T. Wu, M.-H. Yang, and K.-W. Ke, "The design of qos provisioning mechanisms for wireless networks," in *Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops '10)*, pp. 756–759, IEEE, 2010.

[29] J. A. Bergstra and C. A. Middelburg, "Itu-t recommendation g.107 : The e-model, a computational model for use in transmission planning," Technical Report, 2003.

[30] R. G. Cole and J. H. Rosenbluth, "Voice over ip performance monitoring," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 2, pp. 9–24, 2001.

[31] J. Janssen, D. De Vleeschauwer, M. Büchli, and G. H. Petit, "Assessing voice quality in packet-based telephony," *IEEE Internet Computing*, vol. 6, no. 3, pp. 48–56, 2002.

[32] K. Duffy and A. J. Ganesh, "Modeling the impact of buffering on 802.11," *IEEE Communications Letters*, vol. 11, no. 2, pp. 219–221, 2007.

[33] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," *IEEE/ACM Transactions on Networking*, vol. 15, no. 1, pp. 159–172, 2007.

[34] M. Laddomada, F. Mesiti, M. Mondin, and F. Daneshgaran, "On the throughput performance of multirate IEEE 802.11 networks with variable-loaded stations: analysis, modeling, and a novel proportional fairness criterion," *IEEE Transactions on Wireless Communications*, vol. 9, no. 5, pp. 1594–1607, 2010.

[35] K. R. Duffy, "Mean field markov models of wireless local area networks," *Markov Processes and Related Fields*, vol. 16, no. 2, pp. 295–328, 2010.

[36] A. J. Estepa, J. M. Vozmediano, J. López, and R. M. Estepa, "Impact of voip codecs on the energy consumption of portable devices," in *Proceedings of the 6th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks, ser. (PM2HW2N '11)*, pp. 123–130, ACM, New York, NY, USA, 2011.

[37] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proceedings of the Global Communications Conference (GLOBECOM '15)*, pp. 1–6, IEEE, 2015.

[38] N. C. Taher, Y. Ghamri-Doudane, B. El Hassan, and N. Agoulmine, "An efficient model-based admission control algorithm to support voice and video services in 802.11e WLANs," in *Proceedings of the Global Information Infrastructure Symposium (GIIS '09)*, pp. 1–8, IEEE, June 2009.

[39] J. Zhu and A. O. Fapojuwo, "A new call admission control method for providing desired throughput and delay performance in IEEE802.11e wireless LANs," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 701–709, 2007.

[40] D. Pong and T. Moors, "Call admission control for ieee 802.11 contention access mechanism," in *Proceedings of the Global Telecommunications Conference (GLOBECOM '03)*, vol. 1, pp. 174–178, IEEE, 2003.

[41] Q. Zhao, D. H. K. Tsang, and T. Sakurai, "Modeling nonsaturated IEEE 802.11 DCF networks utilizing an arbitrary buffer size," *IEEE Transactions on Mobile Computing*, vol. 10, no. 9, pp. 1248–1263, 2011.

[42] G. Pei and T. R. Henderson, "Validation of ns-3 802.11b phy model," Boeing Research and Technology, Technical Report, 2009.

[43] T. R. H. G. Pei, "Validation of ofdm error rate model in ns-3," Boeing Research and Technology, Technical Report, 2010.

[44] H. Assem, D. Malone, J. Dunne, and P. O'Sullivan, "Monitoring voip call quality using improved simplified e-model," in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC '13)*, pp. 927–931, IEEE, 2013.

[45] R. V. Kulkarni and G. K. Venayagamoorthy, "Bio-inspired algorithms for autonomous deployment and localization of sensor nodes," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 6, pp. 663–675, 2010.

[46] S. Sabino and A. Grilo, "Topology control of unmanned aerial vehicle (UAV) mesh networks: a multi-objective evolutionary algorithm approach," in *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*, pp. 45–50, ACM, 2018.

[47] A. Chipperfield, P. Fleming, H. Pohlheim, and C. Fonseca, "Genetic algorithm toolbox for use with matlab," IEE Colloquium on Applied Control Techniques Using Matlab 14, pp. 10-1–10-4, 1995.

[48] H. E. Aguirre, K. Tanaka, and T. Sugimura, "Cooperative model for genetic operators to improve gas," in *Proceedings of the International Conference on Information Intelligence and Systems*, pp. 98–106, IEEE, 1999.

[49] H. E. Aguirre, K. Tanaka, T. Sugimura, and S. Oshita, "Improved distributed genetic algorithm with cooperative-competitive genetic operators," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 5, pp. 3816–3822, IEEE, 2000.

[50] B. Dengiz, F. Altiparmak, and A. E. Smith, "Local search genetic algorithm for optimal design of reliable networks," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 3, pp. 179–188, 1997.

[51] X. Zhang and L. Duan, "Fast deployment of UAV networks for optimal wireless coverage," *IEEE Transactions on Mobile Computing*, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 2729370, 19 pages
https://doi.org/10.1155/2019/2729370

Research Article

Access Time Analysis of MCPTT Off-Network Mode over LTE

Yishen Sun, Wesley Garey, Richard Rouil, and Priam Varin

Wireless Networks Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Correspondence should be addressed to Yishen Sun; yishen.sun@nist.gov

Received 30 November 2018; Accepted 27 January 2019; Published 2 April 2019

Guest Editor: Maurizio Casoni

Public safety organizations around the world started migrating toward Long-Term Evolution (LTE) networks to support the increasing needs for video and data. To address the unique voice communication requirements of first responders, the 3rd Generation Partnership Project (3GPP) introduced new capabilities that aim at providing similar functionalities as the traditional Land Mobile Radio (LMR) systems, namely, Direct Mode communication and mission critical push-to-talk (MCPTT). Direct Mode communication, also called Proximity Services (ProSe), allows public safety users to communicate directly with each other regardless of the network status. MCPTT was the first mission critical service, and first application, standardized by 3GPP to provide both on- and off-network voice capability. Assessing the performance of those capabilities is critical to accelerate their deployment and adoption by first responders. In this study, we evaluate the performance of an off-network mode MCPTT device over ProSe by focusing on the access time, a measure of the delay incurred before a user can talk. We develop analytical models for various types of calls and verify the accuracy of the predicted access time using ns-3 simulations. We perform sensitivity analysis to show the validity of the models for various scenarios. Finally, we show how the models can be used to guide parameter configuration for both MCPTT and ProSe to optimize the performance.

1. Introduction

Public safety agencies require real-time applications, and voice communication remains the most important means of communication for first responders in emergency situations. While most existing push-to-talk (PTT) communication is carried over Land Mobile Radios (LMRs) to achieve public safety grade voice quality and reliability, first responders have already adopted broadband technology, including 3rd Generation Partnership Project (3GPP) Long-Term Evolution (LTE), to access video and data. Thanks to the effort of public safety organizations around the world, 3GPP has extended LTE's capabilities to support many of the situations faced by first responders and increase the network's reliability. In particular, extensions were made to fulfill the mission critical voice communication requirements defined by the public safety community [1, 2]. Those requirements include PTT, which is the standard means of communicating amongst first responders (similar to walkie-talkie), and Direct Mode, which allows first responders to communicate unit-to-unit without using the network. Also included is the ability to talk to multiple units at once (i.e., group call), the identification of the person speaking, and notification of emergency situations.

PTT is not new, and a lot of research has been done on the Open Mobile Alliance (OMA) PTT over Cellular (PoC) services [3] in recent years. Early evaluations of PoC were done over 3G technologies and the results showed significant setup delays [4] and talking wait time on the order of seconds [5, 6]. Even over LTE, PoC is far from being able to provide the performance needed by first responders. For example, Kuwadekar and Al-Begain evaluated PoC using a real network under different conditions [7]. The call setup delay was measured in the testing, which is the time from when the PTT button was pressed to the time when all the UEs were connected so that the user could speak. The results showed that if the network was not loaded, the call setup delay ranged from 0.7 s to 4.6 s, and that the average call setup delay was 2.556 s, which is higher than what is seen in LMR systems. Kuwadekar and Al-Begain also showed that PTT call setup delay increased roughly linearly with respect to the amount of emulated background calls that they injected

into the network, and the call setup delay exceeded 10 s once the number of simultaneous background calls exceeded 700. Tests of PTT during a festival, when the local network was being used by up to around 10,000 people, resulted in a 100% call failure rate. Those high delays are attributed to the Session Initiation Protocol (SIP) used by PoC and provided by the Internet Protocol (IP) Multimedia Subsystem (IMS). An enhanced version of PoC for public safety, called Push-to-Communicate for Public Safety (PCPS) [8], was developed and aligned with 3GPP efforts to standardize a PTT application capable of meeting public safety needs in terms of features and performance. This Mission Critical Push-to-Talk (MCPTT) application was introduced in LTE Release 13 and further enhanced in later releases [9]. The current efforts to evaluate MCPTT mainly focus on on-network scenarios, where the devices are connected to the network via a base station. In [10], Kim et al. proposed an algorithm to shorten the time of the initial uplink transmission in order to reduce the control plane latency. In [11], Solozobal et al. looked at using mobile edge computing to provide a flexible architecture in 5G networks where the user plane can be deployed closer to the users, thus improving performance. The proposed solution would also allow the network operator to deploy the control plane at the edge of the network to provide a standalone service if the evolved NodeB (eNodeB) is disconnected from the core network.

To support Direct Mode capability in LTE, 3GPP introduced Proximity Services (ProSe) capability in Release 12 and has continued to enhance ProSe in subsequent releases [12]. ProSe allows a piece of User Equipment (UE), such as a laptop or a phone, to discover and communicate with other UEs. ProSe supports both one-to-one and one-to-many transmissions, thus allowing for efficient group communication. The research on ProSe communication has primarily focused on the coverage and capacity aspect of this new feature. Wang and Rouil evaluated the expected device-to-device (D2D) coverage using ProSe [13] for various channel conditions and conducted a sensitivity analysis to determine the main factors impacting coverage. This work considered only a single link and did not capture the effects of multiple users accessing the same resources. Griffith et al. developed analytical and simulation models to evaluate the probability that a control packet reaches a certain number of devices assuming a particular control channel resource pool configuration [14]. In this work, multiple devices were considered but conservative assumptions were made regarding the channel effects. A follow-up work by Griffith et al. was done to evaluate data transmissions over the shared channel [15] and a similar work was completed by Cipriano and Panaitopol [16]. The results of those contributions can help operators allocate resources based on a desired performance target but do not consider end-to-end performance metrics such as latency.

To the best of our knowledge, this is the first work that evaluates the performance of off-network MCPTT over ProSe. Using the MCPTT access time as the key performance indicator (KPI), we propose analytical models to estimate the access time under different scenarios. The models can be used to guide parameter configuration, and to better understand the interaction between MCPTT and ProSe. We have also developed a ns-3 model for simulating MCPTT over ProSe [17], which we used to validate the predicted delays. Major functions of the simulation model include floor control, call control, call type control, and emergency alert. All protocol operations are based on 3GPP MCPTT specifications [18, 19]. We designed test cases and used them to check the correctness of the protocol models that we implemented in the simulator [20].

The rest of the paper is organized as follows. In Section 2, we describe D2D communication over ProSe and our baseline model to estimate the transmission delay. In Section 3, we leverage the ProSe model to develop a set of theoretical models to estimate the MCPTT access time for the various types of calls. We included simulation results to compare with our analytical estimates. In Section 4, we conduct a sensitivity analysis of the ProSe and MCPTT parameters. Section 5 concludes the paper and outlines future work.

2. Modeling of Message Transmission Time over ProSe

2.1. D2D Communication over ProSe Sidelink. The need for wireless communication among public safety users while out-of-coverage can be met by utilizing Mode 2 of D2D communications supported by 3GPP LTE [12]. In Mode 2, UEs operate without any network supervision, which is the case when the UEs are out of coverage of any base stations, and voice and data communications are carried over the air interface referred to as the ProSe sidelink, or "sidelink" for short. D2D communication over sidelink is performed over periodically repeating sets of sidelink resources in the time domain, where a single set of sidelink resources compose a sidelink period. Each sidelink period contains instances of two channels that are separated in the time domain: the Physical Sidelink Control Channel (PSCCH) and the Physical Sidelink Shared Channel (PSSCH), as depicted in Figure 1. Each channel is defined by a resource pool, whose elements are defined by a combination of certain Resource Block (RB) indexes in the frequency domain, and certain subframe indexes in the time domain. The transmitting UE uses the PSCCH to send a Sidelink Control Information (SCI) message to indicate to which UEs its subsequent data messages are addressed, where and when the data transmissions on the PSSCH will occur (i.e., resource assignment in the time and frequency domains), and how (i.e., the modulation and coding scheme). Upon successful reception of the SCI message, any receiving UEs can tune to the corresponding resources in the PSSCH. Transmissions in the PSSCH use a Time Resource Pattern (TRP), which is a subframe indication bitmap of a fixed length N_{TRP} (e.g., 8 subframes for LTE in Frequency Division Duplex (FDD) mode) repeated in the time domain through the length of PSSCH, to identify which subframes are used by a transmitting UE. Each TRP is identified by an index I_{TRP} corresponding to the predefined subframe indication bitmap established in [21]. Every data transmission on the PSSCH is performed with four HARQ transmissions. In addition, if outgoing data is to be scheduled for transmission in a given occurrence of the PSSCH, the data has to be available at least 4 ms before the beginning of the

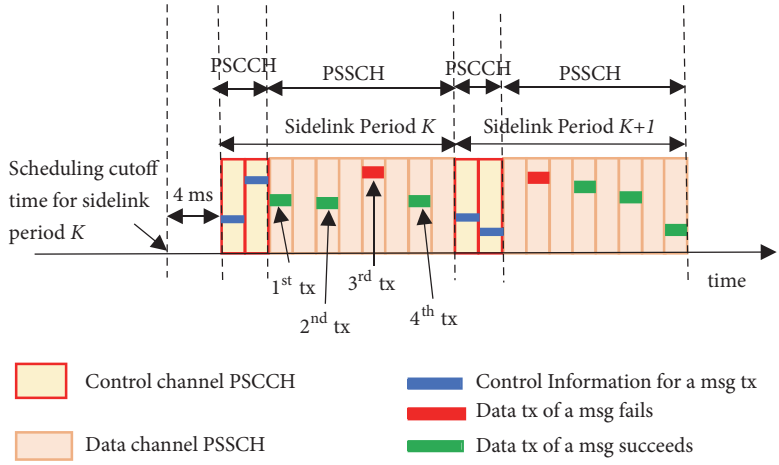


Figure 1: D2D communication over ProSe Sidelink, with periodically repeating control and shared channels.

corresponding sidelink period. In this paper, we identify this gap as the scheduling cutoff time. Sidelink communication is considered half duplex, which means that a device is not able to receive messages if it is transmitting in the same subframe.

2.2. One Way Transmission Time. Let X be a message that a transmitting UE must send and let the one-way transmission time of message X over the sidelink, denoted as $T_{tx}(X)$, be the duration from when the message X becomes available at the transmitting UE to the time when message X is received successfully by the receiving UE. The calculation of $T_{tx}(X)$ considers the packet error rate (PER) of each transmission attempt, the random occurrence of PTT button push timing, and ProSe configuration parameters. To be more specific, $T_{tx}(X)$ can be written as a function $f()$ that has the form:

$$T_{tx}(X) = f\left(T_{SL}, T_{PSCCH}, T_{delay}, [p_1, \ldots, p_N], Prob.(PTT\ pushing), N_{TRP}, I_{TRP}, N, l\right), \quad (1)$$

where

(i) T_{SL} is the duration of a sidelink period;

(ii) T_{PSCCH} is the duration of the control channel (PSCCH) of a sidelink period;

(iii) T_{delay} is the transmission plus propagation delays from the sending UE to the targeted receiving UE;

(iv) $p_n, n = 1, 2, \ldots, N$, is the PER of the n-th transmission attempt within a sidelink period, with N being the total number of transmission attempts of a packet in one sidelink period;

(v) *Prob.* (*PTT pushing*) represents the distribution of the PTT button push timing;

(vi) N_{TRP} and I_{TRP} are ProSe time resource pattern related parameters;

(vii) l is an integer between 1 and L, with $L = \lfloor (T_{SL} - T_{PSCCH})/(N \times N_{TRP}) \rfloor$, and the MCPTT message is included in the l-th Transmission Block (TB) of the sidelink period data channel (PSSCH).

Since $T_{tx}(X)$ remains the same for all MCPTT messages discussed in this paper and thus does not depend on X, we refer to it as T_{tx} for short.

The minimum value of T_{tx} that one message may experience occurs when the following three conditions are met, as shown in Figure 2(a):

(1) the message becomes available for transmission right before the scheduling cutoff time of the upcoming sidelink period K (i.e., 4 ms ahead of the start of sidelink period K);

(2) the message is scheduled to be transmitted in the first subframe of the data channel of the sidelink period (e.g., $I_{TRP} = 0$); and

(3) the first transmission attempt of the message is successfully received by the targeted UE.

Thus, the minimum value for T_{tx} in units of milliseconds is

$$\min(T_{tx}) = 4 + T_{delay} + T_{PSCCH}; \quad (2)$$

The maximum value of T_{tx} that one message may experience, assuming that only the last transmission attempt is successfully received by the targeted UE, is shown in Figure 2(b). The message transmission time can be even larger if none of the transmission attempts in one sidelink period succeed, and we will discuss this situation in Section 4. When both of the following two conditions are met, the maximum value of T_{tx} shown in Figure 2(b) occurs:

(1) the message becomes available for transmission right after the scheduling cutoff time of the upcoming sidelink period K, so it must wait for the full length of the sidelink period K and is scheduled to be transmitted in the subsequent sidelink period $K + 1$;

(2) the last transmission attempt of the message is scheduled to be transmitted in the last subframe of the data channel of the sidelink (e.g., $I_{TRP} = 7$) period.

Thus, the maximum value of T_{tx} in units of milliseconds is

$$\max(T_{tx}) = 3 + T_{delay} + 2 * T_{SL}; \quad (3)$$

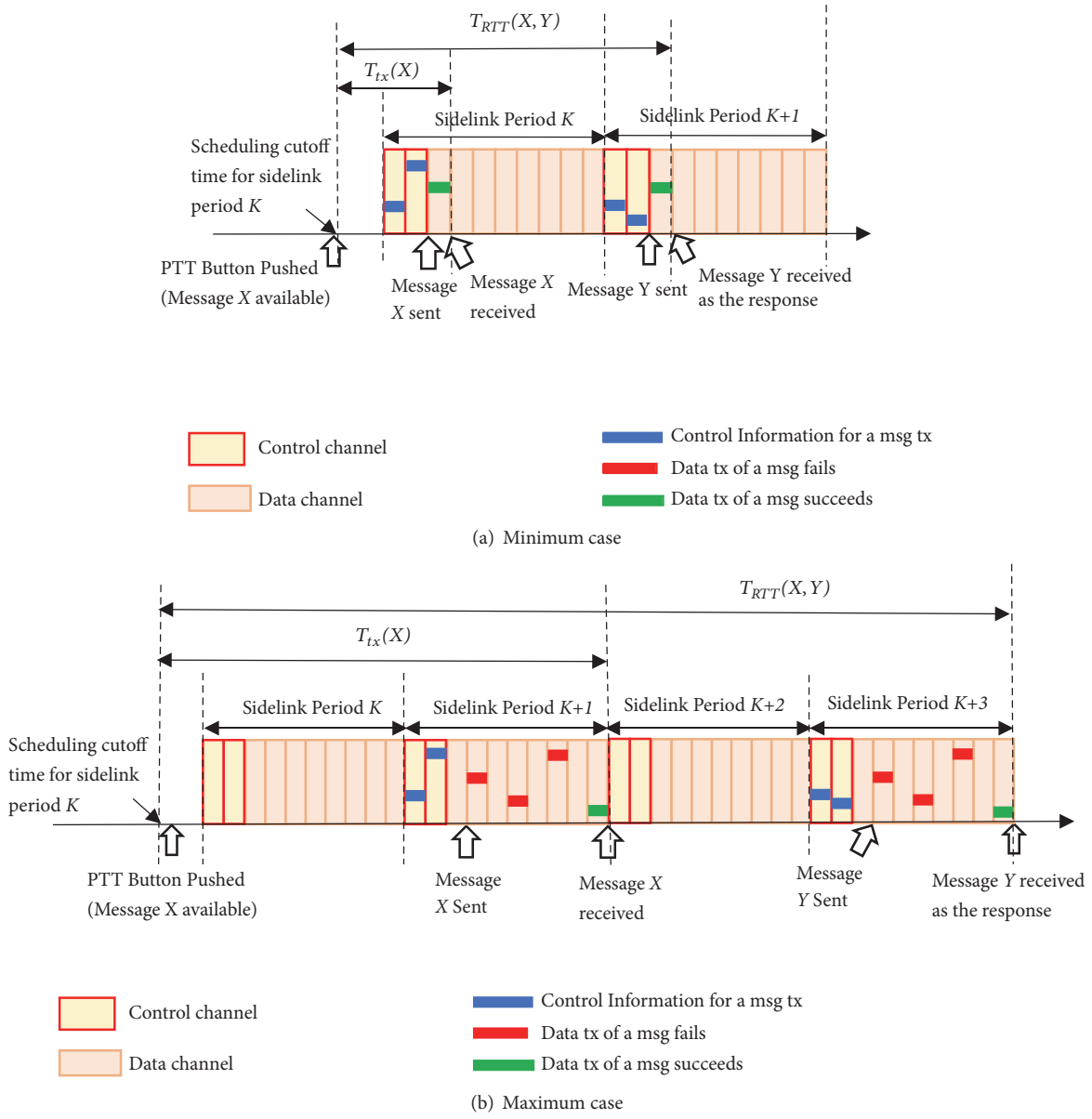


Figure 2: Message TRANSMISSION TIME over ProSe T_{tx} and T_{RTT}.

Assume the l-th TB is the TB in which message X is assigned to be transmitted. Let T_{preSL} denote the time between the moment the PTT button at the transmitting UE is pushed and the beginning of the sidelink period in which the message is transmitted. There are four components of the message transmission time T_{tx}: (1) T_{preSL}; (2) the control channel duration T_{PSCCH}; (3) the time offset of the l-th TB from the beginning of the current instance of the data channel PSSCH which equals $(l-1) \times N_{TRP} \times N$; and (4) the duration from the beginning of the l-th TB to the time that message X is sent out successfully within the l-th TB. Note that for a message to be transmitted with ProSe parameters $I_{TRP} = i$ and assuming the message is transmitted successfully in the n-th attempt initially, component (4) equals $(n-1) \times N_{TRP} + i + T_{delay}$. This is because for each failed transmission attempt, an additional N_{TRP} ms delay will be added, and within each round of transmission attempt, there is i ms time offset introduced for a message transmitted with TRP bitmap parameter $I_{TRP} = i$. Therefore, the average of component (4) shall be calculated by conditioning on the possible I_{TRP} values i and, given i, by additionally conditioning on the index number of the first successful transmission attempt n. Also, since we are looking at the delay associated with a successful message transmission, we use the law of total probability and condition the value of component (4) on the event that the message transmission succeeds in exactly n attempts for $n = 1, 2, \ldots, N$; the probability of this event is $(1-p_n)\prod_{m=0}^{n-1} p_m$. Furthermore, the probability of message transmission failure is the probability of N transmission attempt failures, which is $\prod_{n=1}^{N} p_n$. Thus,

the message transmission success probability is $1 - \prod_{n=1}^{N} p_n$, and the average value of T_{tx} is

$$\begin{aligned} Avg\left(T_{tx}\right) &= E\left(T_{preSL}\right) + T_{PSCCH} + (l-1) \times N_{TRP} \times N + \frac{1}{1-\prod_{n=1}^{N} p_n} \sum_{i=0}^{(N_{TRP}-1)} Prob\left(I_{TRP} = i\right) \\ &\cdot \left[\sum_{n=1}^{N}\left(\left((n-1) \times N_{TRP} + i + T_{delay}\right)\left(1-p_n\right)\prod_{m=0}^{n-1} p_m\right)\right] \\ &= E\left(T_{preSL}\right) + T_{PSCCH} + (l-1) \times N_{TRP} \times N + \sum_{i=0}^{(N_{TRP}-1)} Prob\left(I_{TRP} = i\right) \\ &\cdot \left[\left(i + T_{delay}\right) + \frac{N_{TRP}}{1-\prod_{n=1}^{N} p_n} \sum_{n=1}^{N-1}\left(n\left(1-p_{n+1}\right)\prod_{m=1}^{n} p_m\right)\right] \end{aligned} \tag{4}$$

where $p_0 = 1$.

2.3. Roundtrip Transmission Time. The roundtrip message transmission time over the sidelink, denoted as $T_{RTT}(X, Y)$, is defined as the duration from when message X is available at UE A until UE A receives the response message Y from the corresponding UE (denoted as UE B for simplicity) successfully over the sidelink channel, then

$$\begin{aligned} T_{RTT}\left(X, Y\right) = T_{tx}\left(X\right) + T_B + T_{boundary} \\ + \left(T_{tx}\left(Y\right) - T_{preSL}\left(Y\right)\right) \end{aligned} \tag{5}$$

where T_B is the time between when message X is received by UE B and the time when the response message Y is available at UE B as defined by the protocol (e.g., backoff timers); and $T_{boundary}$ is the time that a message has to wait before the beginning of the sidelink period in which the message is transmitted. This overhead time is due to the sidelink period boundary restriction.

Note that $T_{RTT}(X, Y)$ is not the simple summation of $T_{tx}(X)$, T_B, and $T_{tx}(Y)$. This is because the available timing of message Y is dependent on the available timing of message X and T_B, which is no longer the independent PTT push timing that is assumed for UE B when calculating $T_{tx}(Y)$.

The minimum $T_{RTT}(X, Y)$ that UE A may expect is

$$\begin{aligned} \min\left\{T_{RTT}\left(X, Y\right)\right\} \\ = 4 + \left\lceil \frac{T_{PSCCH} + T_{delay} + \min\left\{T_B\right\} + 4}{T_{SL}} \right\rceil T_{SL} \\ + T_{PSCCH}, \end{aligned} \tag{6}$$

as shown in Figure 2(a) by assuming $T_B = 0$.

As can be seen from (6), $\min\{T_{RTT}(X, Y)\}$ and thus $T_{RTT}(X, Y)$ is always greater than T_{SL}.

The maximum $T_{RTT}(X, Y)$ that UE A may experience is

$$\begin{aligned} \max\left\{T_{RTT}\left(X, Y\right)\right\} = 3 \\ + \left(1 + \left\lceil \frac{T_{PSCCH} + N \times N_{TRP} + T_{delay} + \max\left\{T_B\right\} + 3}{T_{SL}} \right\rceil\right) \\ \cdot T_{SL} + T_{PSCCH} + N \times N_{TRP} + T_{delay} \end{aligned} \tag{7}$$

as shown in Figure 2(b) by assuming $T_B = 0$.

The calculation of average $T_{RTT}(X, Y)$ starts from rearranging (5) into mutually indepenedent components:

$$\begin{aligned} Avg\left(T_{RTT}\left(X, Y\right)\right) &= E\left(T_{tx}\left(X\right) + T_B + T_{boundary} + \left(T_{tx}\left(Y\right) - T_{preSL}\left(Y\right)\right)\right) \\ &= E\left(T_{preSL}\left(X\right) + \left(T_{tx}\left(X\right) - T_{preSL}\left(X\right)\right) + T_B + T_{boundary} + \left(T_{tx}\left(Y\right) - T_{preSL}\left(Y\right)\right)\right) \\ &= E\left(T_{tx}\left(X\right) - T_{preSL}\left(X\right) + T_B + T_{boundary}\right) + E\left(T_{tx}\left(Y\right)\right) \\ &= T_{SL} * E\left(K_{RTT}\right) + E\left(T_{tx}\left(Y\right)\right) \end{aligned} \tag{8}$$

where K_{RTT} is one less than the number of sidelink periods from the sidelink period in which message X is transmitted successfully to the sidelink period in which message Y is transmitted successfully, so $K_{RTT} = (T_{tx}(X) - T_{preSL}(X) + T_B + T_{boundary})/T_{SL}$. Similar to component (4) in (4), the average of K_{RTT} shall

be calculated considering both possible I_{TRP} values i and possible numbering of the first successful transmission attempts n. In addition, the randomness introduced by T_B needs to be included as well.

$$
\begin{aligned}
E\left(K_{RTT}\right) &= \left[\sum_{n=1}^{N}\left(\left(\left(1-p_n\right)\prod_{m=0}^{n-1}p_m\right)\sum_{k=1}^{K_{max}(n)}\left(k * Prob\left(K_{RTT}=k \mid\right.\right.\right.\right. \\
&\left.\left.\left.\left. \text{the first successful transmission is attempt } n\right)\right)\right)\right] \times \frac{1}{1-\prod_{n=1}^{N}p_n} \\
&= \frac{1}{1-\prod_{n=1}^{N}p_n}\left\{\sum_{n=1}^{N}\left[\left(\left(1-p_n\right)\prod_{m=0}^{n-1}p_m\right)\sum_{k=1}^{K_{max}(n)}k\sum_{i=0}^{(N_{TRP}-1)}Prob\left(I_{TRP}=i\right)\right.\right. \\
&\left.\left.\times Prob\left(\left(k-1\right)T_{SL} \le \left(T_{PSCCH}+4+\left(n-1\right)\times N_{TRP}+\mathrm{i}+T_{delay}+T_B < k * T_{SL}\right)\right)\right]\right\}
\end{aligned}
\tag{9}
$$

where $p_0 = 1$; and $K_{max}(n) = \lceil (T_{PSCCH} + n \times N_{TRP} - 1 + T_{delay} + \max\{T_B\})/T_{SL} \rceil$ is the maximum possible K_{RTT} for a given n. By combining (8) and (9), the average value of $T_{RTT}(X, Y)$ is

$$
\begin{aligned}
&Avg\left(T_{RTT}\left(X,Y\right)\right) \\
&= \left[\sum_{n=1}^{N}\left(\left(\left(1-p_n\right)\prod_{m=0}^{n-1}p_m\right)\sum_{k=1}^{K_{max}(n)}k\sum_{i=0}^{(N_{TRP}-1)}Prob\left(I_{TRP}=i\right)\times Prob\left(T_{BL}\left(n,k,i\right)\le T_B < T_{BU}\left(n,k,i\right)\right)\right)\right] \\
&\times \frac{T_{SL}}{1-\prod_{n=1}^{N}p_n} + Avg\left(T_{tx}\right)
\end{aligned}
\tag{10}
$$

where $Avg(T_{tx})$ is calculated by (4);

$$
\begin{aligned}
T_{BL}\left(n,k,i\right) &= \left(k-1\right)T_{SL} - T_{PSCCH} - 4 \\
&\quad - \left(n-1\right)N_{TRP} - i - T_{delay}; \\
\text{and } T_{BU}\left(n,k,i\right) &= k * T_{SL} - T_{PSCCH} - 4 \\
&\quad - \left(n-1\right)N_{TRP} - i - T_{delay}.
\end{aligned}
\tag{11}
$$

2.4. Baseline Configuration Parameters for Evaluation. Analytical estimates and simulation results presented in Sections 3 and 4 assume the baseline configuration shown in Table 1 unless otherwise noted. Note that the PER of a single transmission attempt is set at 0.1 at most, since 0.1 is the common design assumption of LTE PER when Hybrid automatic repeat request (HARQ) Acknowledegment (ACK)/Negative Acknowledgement (NACK) is in use. Given that there is no HARQ ACK/NACK over the ProSe PSSCH meaning that all four HARQ transmissions always occur, the PER should not be set to a more aggressive value higher than 0.1.

We assume in the simulations that the timing of a PTT button push is uniformly distributed within a sidelink period T_{SL}, and $l = 1$ (i.e., the message is transmitted in the first TB of the sidelink, which is reasonable given the importance of public safety related communication). It is also assumed that ProSe TRP parameters are configured as $k_{TRP} = 1$ and $N_{TRP} = 8$. This means that only one of the 8 subframes in a TRP will be used to send data and $Prob(I_{TRP} = i) = 1/N_{TRP} = 1/8$. Then (4) is simplified as

$$
\begin{aligned}
Avg\left(T_{tx}\right) &= 7.5 + \frac{T_{SL}}{2} + T_{PSCCH} + T_{delay} \\
&\quad + \frac{8\left(p_1 + p_1p_2 + p_1p_2p_3 - 3p_1p_2p_3p_4\right)}{1-p_1p_2p_3p_4}
\end{aligned}
\tag{12}
$$

in units of ms.

We also assume that the MCPTT user only needs to push the PTT button to initiate the request to speak and is not required to follow up the request with any further actions, such as acknowledgement. In addition, since there are some variations in MCPTT message sizes (e.g., call announcement message is long) and the size of resource allocated to each UE (e.g., the resource allocated for a voice message may be small to start with), our modeling assumes that MCPTT messages are not piggybacked, in order to get the conservative

Table 1: Baseline configuration parameters used in the MCPTT access time evaluation.

Parameter Name	Symbol	Value
Sidelink period (ms)	T_{SL}	40, 60, 80, 120, 160, 240, 280, 320
PSCCH duration (ms)	T_{PSCCH}	8
# transmission attempts in one sidelink period	N	4
Time Resource Pattern parameters	N_{TRP}	8
	k_{TRP}	1
PER of the 4 transmission attempts in each sidelink period	PER	[0.1, 0.1, 0.1, 0]
Transmission + Propagation delays (ms)	T_{delay}	0.93

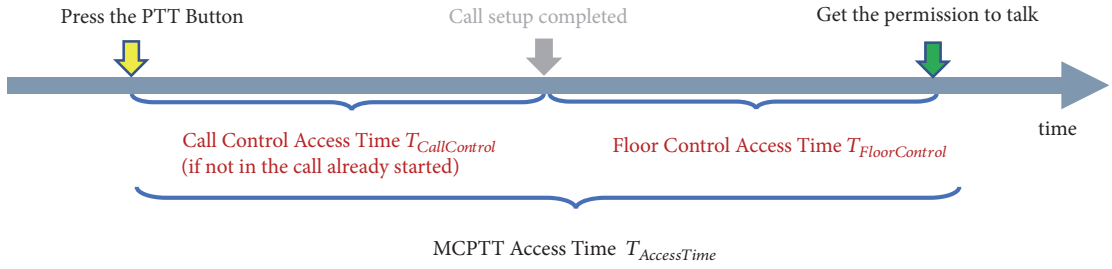


Figure 3: Components of MCPTT access time.

estimates/results. When piggybacking is allowed, the access time may be further reduced.

3. Modeling of MCPTT Access Time

3.1. Overview. According to 3GPP LTE specifications, "The MCPTT access time is defined as the time between when an MCPTT User requests to speak (normally by pressing the MCPTT control on the MCPTT UE) and when this user gets a signal to start speaking." [9]. Based on protocol details, the MCPTT access time ($T_{AccessTime}$) can be further divided into two components as illustrated in Figure 3: call control access time ($T_{CallControl}$) and floor control access time ($T_{FloorControl}$). Call Control Access Time ($T_{CallControl}$) is the time from when the MCPTT user presses the PTT button to when the MCPTT UE establishes a call successfully. This component occurs only when the call which the MCPTT UE wants to join/start does not exist yet. Floor control access time ($T_{FloorControl}$) is the time from when the MCPTT UE establishes a call successfully (if not already in a call) or when the MCPTT user presses the PTT button (if in a call already), whichever is later, to when the MCPTT UE becomes the floor arbitrator. In MCPTT off-network mode group call operation, the floor arbitrator is the current speaker who is the only group member that is allowed to speak to the group and that also arbitrates the floor in case other members of the group call (i.e., other floor participants) request that they be granted the floor so that they can talk.

MCPTT off-network mode supports the following three categories of calls: basic group call, broadcast group call, and private call. The analytical modeling of MCPTT access time associated with basic group call is analyzed first in Section 3.2, because the call control procedures and floor control procedures of the basic group call are very comprehensive. Then the analytical modeling of MCPTT access time associated with broadcast group call and private calls is presented in Sections 3.3 and 3.4, respectively, which reuses a subset of the modeling of basic group call procedures. We obtain simulation results assuming two users for all three call categories and compared these results with their analytical estimate counterparts in the corresponding sections. Simulation results presented throughout this paper are generated by running the ns-3 model for MCPTT over ProSe [17], especially the call control module and floor control module. Parameters configuration and simulation setup follow the description in Section 2.4, unless otherwise noted.

3.2. Basic Group Call. From the viewpoint of MCPTT access time performance evaluation, the various situations that an MCPTT client faces when pushing the PTT button to talk in a basic group call (with group ID *nn*) can always be categorized into one of the five scenarios summarized in Table 2.

In Sections 3.2.1 and 3.2.2, we explain the call control and the floor control access procedures, and access time models are described. In Section 3.3 we derive the total access time $T_{AccessTime}$ experienced by an MCPTT client under Scenarios A through E from different combinations of $T_{CallControl}$ and $T_{FloorControl}$. Note that the value of $T_{FloorControl}$ may be 0 for certain scenario(s), as will be analyzed in Section 3.2.2.

3.2.1. Call Control Access Time $T_{CallControl}$. When an MCPTT client (UE A) wants to join/establish a basic group call with MCPTT group ID *nn* by pushing the PTT button, it has to go through either procedure "*C1: Initiate a new group call*" or procedure "*C2: Join an existing group call,*" which are shown in Figure 4. In the figure, the colors of the message name and the corresponding arrows are consistent with the color of the sending UE.

The call control access procedure we defined above has incorporated the call probe procedure, call setup procedure,

Table 2: Scenarios for basic group call.

Scenario	Group call *nn* exists?	UE A in group call *nn* already?	The floor arbitrator of group call *nn* exists?	$T_{AccessTime}$
A	No	No	No	$T_{CallControl} + T_{FloorControl}$
B	Yes	No	Yes	$T_{CallControl} + T_{FloorControl}$
C	Yes	No	No	$T_{CallControl} + T_{FloorControl}$
D	Yes	Yes	Yes	$T_{FloorControl}$
E	Yes	Yes	No	$T_{FloorControl}$

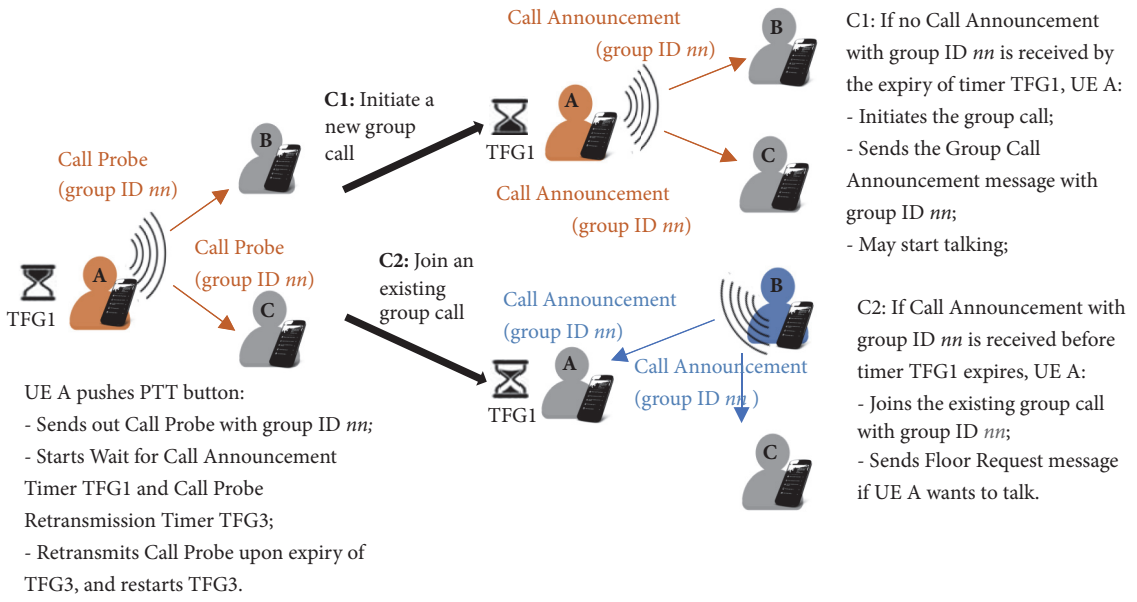


Figure 4: Call control access procedures.

and periodic group call announcement procedure specified in [18]. The decision to go through procedures C1 or C2 is made by UE A (the UE with the PTT button pushed) itself based on its own observation and may not be the correct decision. To be more specific, UE A always chooses C1 under Scenario A, which is the right decision. However, under Scenarios B or C, UE A might make the correct decision, C2, or the wrong decision, C1, depending on how soon UE A hears back from other UEs with respect to the configuration of timer TFG1.

When UE A executes the C1 procedure,

$$T_{CallControl}(C1) = TFG1; \tag{13}$$

where $TFG1$ is the value of "wait for call announcement" timer $TFG1$.

When UE A executes the C2 procedure,

$$T_{CallControl}(C2) = T_{RTT}(\text{"}Call\ Probe\text{"}, \text{"}Call\ Announcement\text{"}); \tag{14}$$

where T_{RTT} is given by (5) and where messages X and Y are the Call Probe and Call Announcement messages, respectively. In (5), $T_B = TFG2_{probe}(M)$, where M is the number of MCPTT UEs in the group call capable of sending out a Call Announcement message in response to receiving a Call Probe message from another UE. Here $TFG2_{probe}(M)$ reflects the time that it takes for at least one UE of the group to respond to a Call Announcement message after receiving the Call Probe message from UE A. It is a random variable that is the minimum of M random variables, each of which is uniformly distributed between 0 seconds and (1/12) seconds (83.3 ms). Therefore,

$$TFG2_{probe}(M) = \begin{cases} 0, & min \\ \dfrac{1000}{12(M+1)}, & avg \\ \dfrac{1000}{12}, & max \end{cases} \tag{15}$$

The minimum, maximum, and average of $T_{CallControl}(C2)$ can be obtained by plugging configuration values into (6), (7), and (10) accordingly.

The simulation study of call control access time in this section focuses on Scenarios B and C, i.e., Procedure C2. This is because there is no uncertainty in Scenario A (i.e., Procedure C1), since $T_{CallControl}$ always equals $TFG1$ when the MCPTT UE establishes a new group call.

Figure 5 shows that, for all sidelink periods T_{SL} defined by 3GPP for LTE ProSe, the analytical model provides good estimates of the minimum, average, and maximum for the

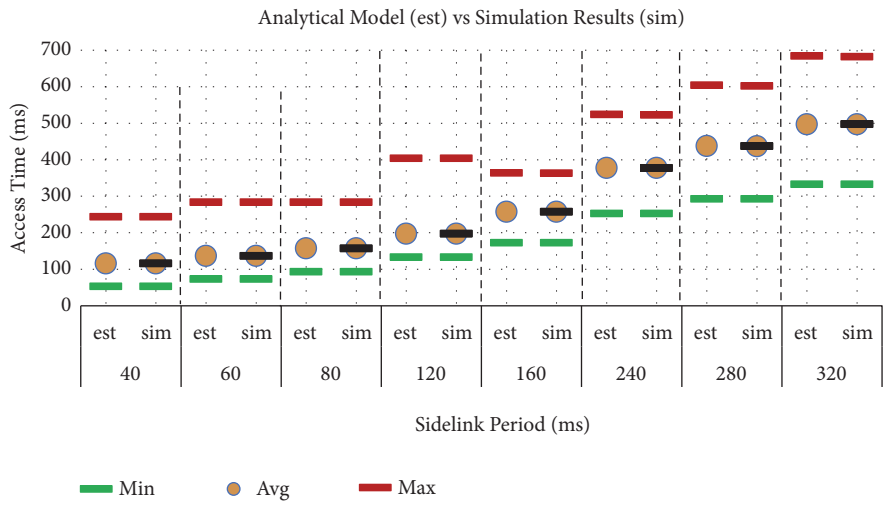


FIGURE 5: Call control access time for joining an existing group call with large value for *TFG*1 with 95% Confidence Intervals for the average values from the simulations.

call control access time experienced by an MCPTT client who goes through procedure C2. The 95% Confidence Interval for the average values obtained from simulation is also shown and is very tight. The minimum and maximum values are observed over all simulations and therefore do not have a Confidence Interval. This is true for all the figures shown in this document. As can be seen from the figure, $T_{CallControl}(C2)$ increases as T_{SL} increases and ranges from as small as 52 ms to more than 680 ms. The *TFG*1 timer value in the simulation is configured to max (1 sec, $6 * T_{SL}$) and the *TFG*3 timer value is configured to $6 * T_{SL}$; thus UE A will make the correct decision because neither timer will expire prematurely.

3.2.2. Floor Control Access Time $T_{FloorControl}$. When an MCPTT client wants to talk and thus pushes the PTT button on UE A, the UE will go through one of the "*F0: new call,*" "*F1: No Response,*" or "*F2: Response Received*" procedures depending on the feedback that it receives from other UEs.

After the MCPTT client (UE A) establishes a group call (i.e., C1 procedure), UE A sends out a Floor Granted message and may start talking, so the corresponding floor control access procedure is as follows.

F0. Send out Floor Granted message to grant the floor to oneself.

After an MCPTT client (UE A) joins an existing group call (i.e., C2 procedure), or if UE A is already a part of an ongoing group call, UE A sends out a Floor Request message. Depending on whether a response is received within a configured time window, UE A may perform the F1 or F2 procedure as illustrated in Figure 6. In the figure, the colors of the message name and the corresponding arrows are consistent with the color of the sending UE. Depending on the relative priority of the floor request and the queueing capability of the current floor arbitrator, procedure F2 may be further divided into three cases:

F2.1 (preemptive floor request). UE A's floor request has preemptive priority, so UE B must grant the floor to UE A immediately;

F2.2 (floor request denied). UE B is busy and does not support queueing requests, so UE B denies the floor to UE A;

F2.3 (floor request queued). UE B is busy but supports queueing of requests, so UE A's floor request is placed in queue and will be handled later.

It is worth pointing out that the time it takes for UE A to receive a response is the same for F2.1, F2.2, and F2.3. However, the 3GPP definition of MCPTT access time focuses on the case of "when this user gets a signal to start speaking" [2]. Therefore, in the following analytical model and simulation study, among the three cases of procedure F2, we focus on procedure "F2.1: preemptive floor request". It is because procedure F2.2 or F2.3 will not grant the floor to UE A, and thus UE A's access procedures are not successful. The model and results can be easily generalized if the MCPTT access time definition is rewritten as "when this user gets a signal whether he/she can start speaking".

The floor control access procedure we defined above has incorporated several state transition procedures specified in [19]. The decision to go through procedure F0, F1, or F2 is made by UE A (the UE with the pushed PTT button) based on its own observation, and this decision may or may not be the correct one. To be more specific, UE A always chooses F1 under Scenario A, which is the right decision. However, under Scenarios B or D, UE A might make the correct decision (F2) or the wrong decision (F1), depending the time until UE A receives a response message from the floor arbitrator UE B and the configuration values of the *T*201 timer and *C*201 counter. The floor control access time of each of the three procedures are as follows:

When F0 procedure is executed,

$$T_{FloorControl}(F0) = 0; \tag{16}$$

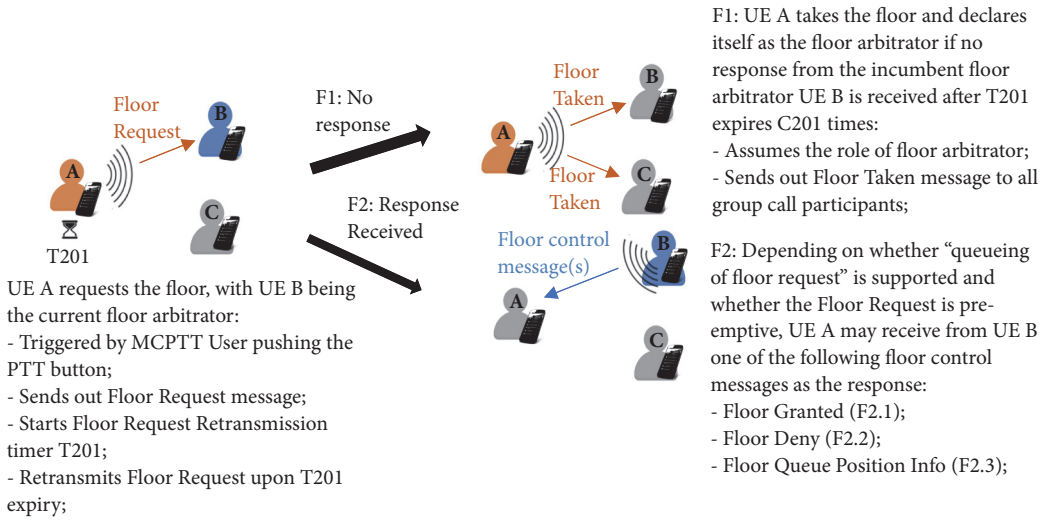


Figure 6: Floor control access procedure.

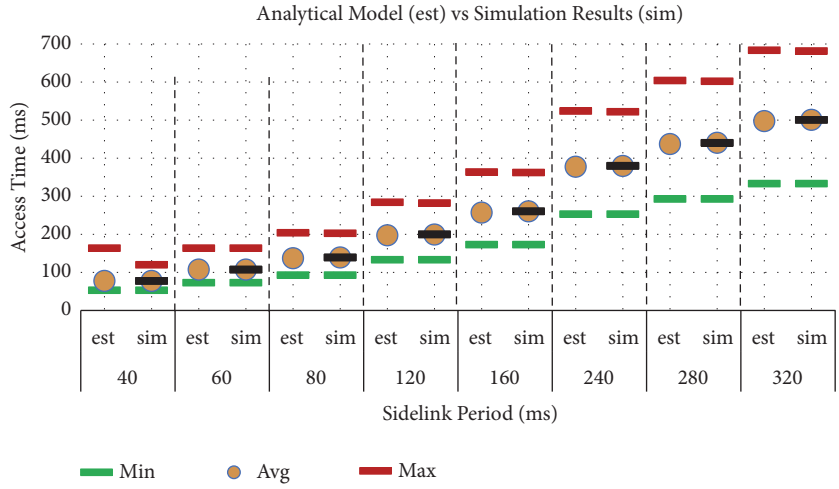


Figure 7: Floor control access time for preemptive floor request with 95% Confidence Intervals for the average values from the simulations.

When F1 procedure is executed,

$$T_{FloorControl}(F1) = C201 * T201; \quad (17)$$

where $T201$ is the "Floor Request Retransmission" timer and $C201$ is the "Floor Request Retransmission" counter. The 3GPP default settings of these two parameters are T_{SL} and 3, respectively.

When the F2.1 procedure is executed,

$$T_{FloorControl}(F2.1) = T_{RTT}(\text{"Floor Request"}, \text{"Floor Granted"}); \quad (18)$$

again using (5), where $T_B = 0$. The minimum, maximum, and average of $T_{FloorControl}(F2.1)$ can be obtained by plugging configuration values into (6), (7), and (10) accordingly.

The simulation study of floor control access time in this section focuses on either Scenario B or D, i.e., procedure F2.1. This is because there is no uncertainty associated with $T_{FloorControl}$ for Scenarios A, C or E: $T_{FloorControl} = C201 * T201$ for Scenarios C & E (i.e., procedure F1), and $T_{FloorControl} = 0$ for Scenario A (i.e., procedure F0). Also, we focus on the case when UE A's floor request is of preemptive priority, i.e., F2.1, as explained in Section 2.4. In the simulation, it is assumed that the incumbent floor arbitrator UE B does not generate Real-time Transport Protocol (RTP) packets during UE A's access time. This assumption will not impact UE A's access time when UE A makes the right decision. We make this assumption to get a conservative estimate of the likelihood of UE A's making the wrong decision, which leads to an undesirable situation where there are multiple floor arbitrators. More detailed definition and analysis of the multiple floor arbitrator situation are given in Section 4.3.

Figure 7 shows that for all sidelink period lengths T_{SL} defined by LTE, the analytical model provides good estimates of the minimum, average and maximum floor control access times experienced by an MCPTT client who goes through procedure F2.1. The $T201$ timer value in the simulation is

Table 3: MCPTT access time $T_{AccessTime}$ for different scenarios.

Scenario	Correct decision	Wrong decision
A	$T_{CallControl}(C1) + T_{FloorControl}(F0)$	N/A
B	$T_{CallControl}(C2) + T_{FloorControl}(F2)$	$T_{CallControl}(C2) + T_{FloorControl}(F1)$ or $T_{CallControl}(C1) + T_{FloorControl}(F0)$
C	$T_{CallControl}(C2) + T_{FloorControl}(F1)$	$T_{CallControl}(C1) + T_{FloorControl}(F0)$
D	$T_{FloorControl}(F2)$	$T_{FloorControl}(F1)$
E	$T_{FloorControl}(F1)$	N/A

configured to $6 * T_{SL}$, so that UE A will make the correct decision.

Comparing Figure 7 with Figure 5, it shows that although the minimum call control access time $T_{CallControl}(C2)$ and floor control access time $T_{FloorControl}(F2.1)$ are very similar, the average and maximum values of $T_{FloorControl}(F2.1)$ are smaller than those of $T_{CallControl}(C2)$ for short sidelink periods (small T_{SL} values). This is because of the extra delay introduced by timer $TFG2$ to the round-trip time during the call control procedure. When T_{SL} is large, however, the impact of $TFG2$ becomes negligible.

3.2.3. Group Call MCPTT Access Time $T_{AccessTime}$. The analytical model of the overall access times for scenarios A through E can be obtained by the proper combination of the analytical models of $T_{CallControl}$ and $T_{FloorControl}$ derived in previous sections, as shown in Table 3. It is worth pointing out that for Scenario A, $T_{FloorControl}(F0)$ is 0 as was explained in Section 3.2.2.

It is worth pointing out that, for Scenarios B, C, and D, UE A may make the correct decision or the wrong decision, depending on the MCPTT parameters' configuration and the one-way or roundtrip message transmission time over the sidelink. When UE A makes the correct decision, the analytical models presented in Section 3 align quite well with the MCPTT access time observed in our simulations. When UE A makes wrong decisions, the simulation results deviate from the analytical estimates, which we analyze in Section 4, together with the discussion of potential negative impacts from incorrect decisions.

3.3. Broadcast Group Call. A broadcast group call is a special group call where the initiating MCPTT user expects no response from the other MCPTT users, so that when the initiator's transmission is complete, so is the call. Therefore, the UE that initiated the broadcast group call is the only MCPTT UE acting as the floor arbitrator. Consequently, only one scenario needs to be evaluated to examine the access time required to establish a new broadcast group call:

Scenario F. UE A establishes a new broadcast group call by pushing the PTT button.

The value of $T_{CallControl}$ under Scenario F is zero, because the UE can initiate a broadcast group call without waiting for timer expiration or messages from other UEs. Similarly, the value of $T_{FloorControl}$ is also zero, because the UE that initiates a broadcast group call becomes the floor arbitrator immediately. Consequently, the overall access time is:

$$T_{AccessTime} = 0; \tag{19}$$

3.4. Private Call. A private call takes place between two MCPTT UEs. For the private call scenarios, a MCPTT UE is in one of two possible states: it is in a private call already or needs to start a new private call; a UE cannot join an existing private call. Consequently, there are only two scenarios that need to be evaluated to obtain the access time of a private call:

Scenario G. UE A establishes a new private call by pushing the PTT button.

Scenario H. UE A is already in a private call and wants to talk by pushing the PTT button.

For Scenario G, the possible call control procedures that UE A may experience are depicted in Figure 8. In the figure, the colors of the message name and the corresponding arrows are consistent with the color of the sending UE. To establish a private call successfully, procedure CP2.1 is the only path, and then UE A becomes the floor arbitrator automatically. Therefore, the overall access time for Scenario G is

$$\begin{aligned} T_{AccessTime} &= T_{CallControl}(CP2.1) \\ &= T_{RTT}(\text{“}Private\ Call\ Setup\ Request\text{”}, \\ &\quad \text{“}Private\ Call\ Accept\text{”}); \end{aligned} \tag{20}$$

which uses (5) to evaluate T_{RTT}. The minimum, maximum, and average of $T_{CallControl}(CP2.1)$ can be obtained by plugging configuration values into (6), (7), and (10), respectively, with $T_B = 0$

Under Scenario G, when the time for UE A to receive a response is greater than $TFP1 * CFP1$, UE A will make the conclusion that the other UE (UE B) cannot be reached and stop trying to establish the call, i.e., procedure CP1.

For Scenario H, the floor control procedure that UE A may experience is very similar to the F2.1 procedure, and the overall access time for Scenario H is from (18)

$$\begin{aligned} T_{AccessTime} &= T_{FloorControl}(F2.1) \\ &= T_{RTT}(\text{“}Floor\ Request\text{”}, \text{“}Floor\ Granted\text{”}); \end{aligned} \tag{21}$$

The simulation study of private call access time in this section focuses on Scenario G, i.e., procedure CP2.1. This

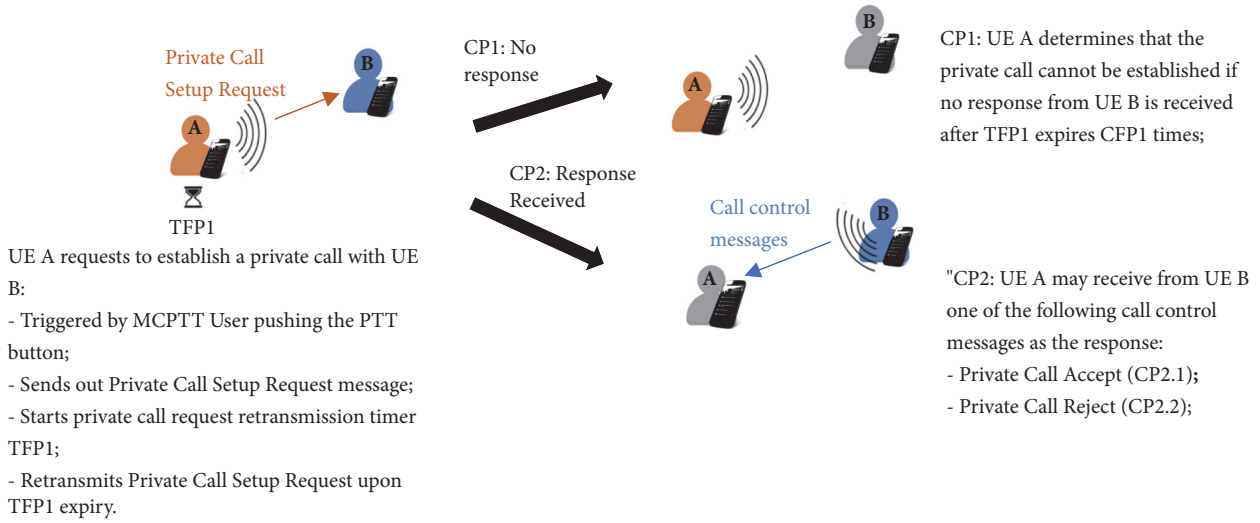


Figure 8: Call control access procedure of private call.

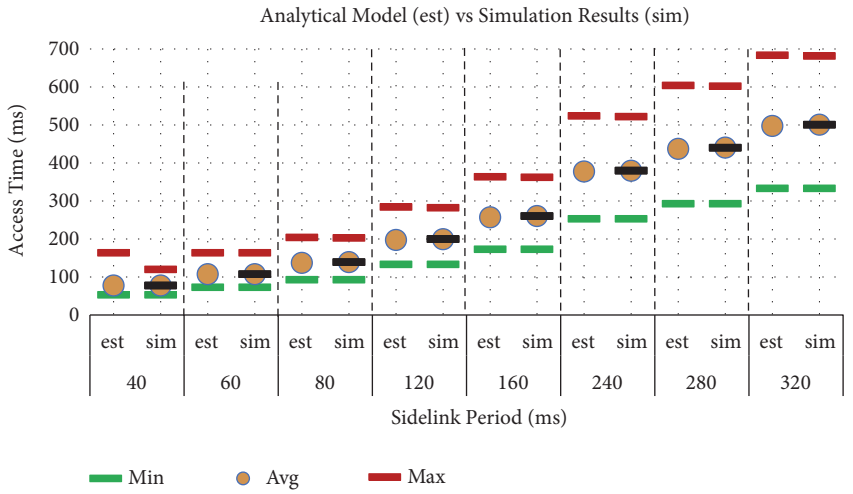


Figure 9: Access time of private call with 95% Confidence Intervals for the average values from the simulations.

is because the access time of Scenario H can use the results for F2.1 presented in Section 3.2.2. Figure 9 shows that, for all sidelink period lengths T_{SL} defined by 3GPP, the analytical model provides good estimates of the minimum, average, and maximum for access times experienced by an MCPTT client going through procedure CP2.1. We observe that the access time and the spread between the minimum and the maximum values increases as T_{SL} increases, for both analytical estimates and simulation results.

4. Impact of MCPTT Parameters on MCPTT Access Time

As already mentioned briefly in Section 3, several important observations can be made based on the analytical and simulation modeling of MCPTT access time.

Observation 1. An MCPTT UE needs more than one sidelink period to send a message and get the response message back over the sidelink.

As can be seen from (6), $\min\{T_{RTT}(X,Y)\}$ and thus $T_{RTT}(X,Y)$ is always longer than T_{SL}. However, the default 3GPP settings of several MCPTT retransmission timers, e.g., $TFG3$ and $T201$, is one T_{SL} only. Such a small timer value may lead to unnecessary retransmissions of certain MCPTT messages and may have negative impacts as illustrated in Sections 4.1 and 4.3.

Observation 2. An MCPTT UE may make wrong decisions about the existence of call and/or floor arbitrators, under some scenarios and with certain parameter configurations.

As summarized in Table 3, UE A's decision may go wrong for Scenarios B, C, and D. These wrong decisions may lead to undesirable situations such as a multiple calls situation or a

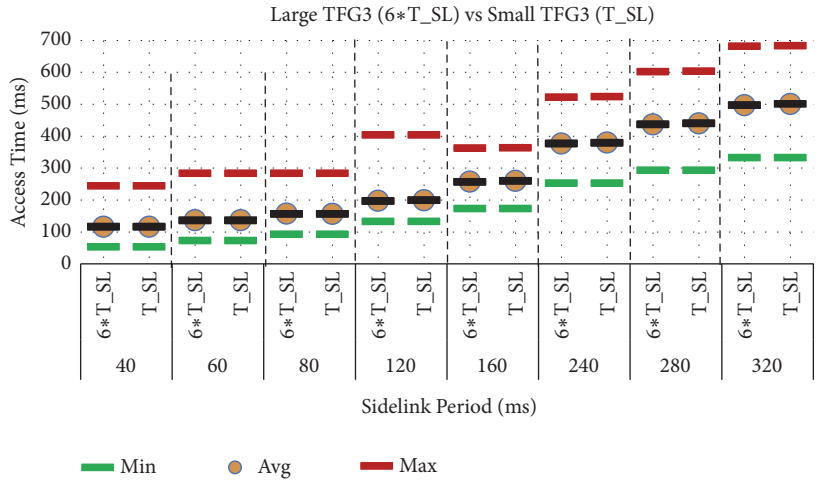


Figure 10: Impact of *TFG3* Timer on call control access time with 95% Confidence Intervals for the average values from the simulations.

Table 4: Impact of *TFG3* configuration.

	T_{SL} (ms)	40	60	80	120	160	240	280	320
Ratio of multiple occupancy to single occupancy	$TFG3 = T_{SL}$	0.510	0.667	0.670	0.576	0.573	0.524	0.530	0.534
	$TFG3 = 6 * T_{SL}$	0	0	0	0	0	0	0	0
Fraction of samples discarded	$TFG3 = T_{SL}$	0.132	0.140	0.092	0.059	0.052	0.024	0.025	0.026
	$TFG3 = 6 * T_{SL}$	0	0	0	0	0	0	0	0

multiple floor arbitrators situation, both of which are further analyzed in Sections 4.2 and 4.3, respectively.

Motivated by the above observations, in the remainder of this section, we study the impact of various MCPTT parameters to evaluate the benefits and drawbacks of different configurations on MCPTT access time performance.

4.1. Impact of "Call Probe Retransmission" Timer TFG3. To evaluate the impact of *TFG3*, we look at the call control access time for Scenarios B and C (group calls). Figure 10 shows the results using the same assumptions as for Figure 5 but with *TFG3* configured at the 3GPP default setting T_{SL} instead of $6 * T_{SL}$.

It appears that varying *TFG3* from T_{SL} to $6 * T_{SL}$ produces no significant difference in the call control access time experienced by an MCPTT client. To further investigate the impact of *TFG3*, we define two other performance metrics called "multiple occupancy" and "single occupancy" to capture the impact of the half duplex effect, which can lead to packet losses at the receiving UE. "Multiple occupancy" refers to a situation in which multiple UEs try to send messages over the air during the same sidelink period, while "single occupancy" refers to a situation in which only one UE tries to send messages over the air during a sidelink period. The half-duplex effect will not occur with "single occupancy" but may occur with "multiple occupancy". As shown in Table 4, the negative impact of *TFG3*'s configuration value's being too small can be seen in the higher ratio of "multiple occupancy" vs "single occupancy," which may lead to more frequent occurrences of the half duplex effect. This is mainly due to unnecessary retransmissions of the "Call Probe" message. The negative impact of *TFG3*'s being too small is also shown through the nonzero ratio of samples discarded. Note that a sample is discarded if both SCI message transmissions in the control channel (PSCCH) fail or all four HARQ transmissions in the data channel (PSSCH) fail, due to either half-duplex effect or channel PER. Given PER = [0.1, 0.1, 0.1, 0] as a simulation assumption, all samples discarded are always the consequence of the half duplex effect. The fraction of samples discarded is dependent on the size of the resource pools allocated to sidelink communication. The positive impact of a small *TFG3* is that the "Call Probe" message may be retransmitted sooner if it is lost.

It is worth pointing out that the occurrence of "multiple occupancy" and thus the half-duplex effect are expected to be more frequent when the number of participating UEs in the group call increases. Therefore, although we do not see any noticeable difference in access time when changing *TFG3* value in the 2-UE system, the impact of *TFG3* will become more obvious for multiple-UE systems. In addition, the occurrence of the half-duplex effect can be reduced by adapting resource pool configuration to the number of participating UEs (e.g., extending the pool in the time domain), which is an interesting research topic by itself.

4.2. Impact of "Wait for Call Announcement" Timer TFG1. When we run the simulation with the same assumptions as for Figure 5 but with *TFG1* configured at the 3GPP default setting of 150 ms instead of $\max(1s, 6 * T_{SL})$, we can see from

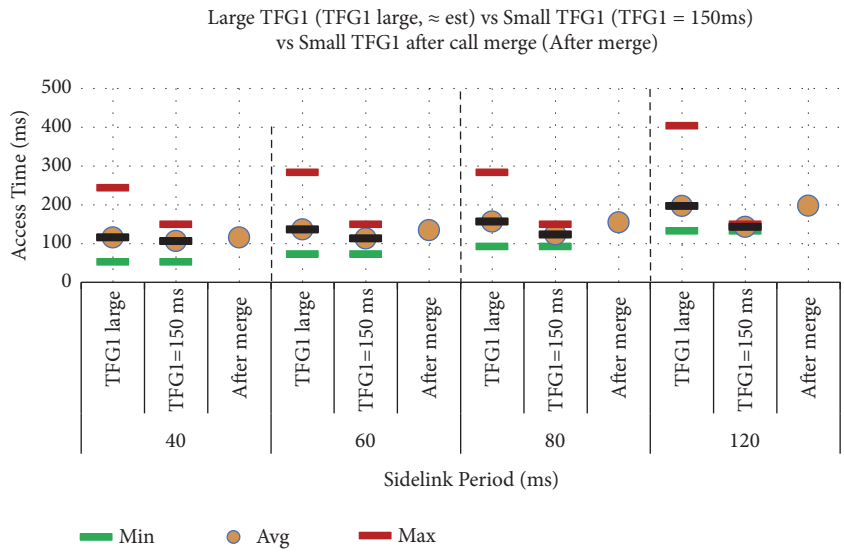


Figure 11: Impact of TFG1 on call control access time with 95% Confidence Intervals for the average values from the simulations.

Figure 11 that the maximum access times experienced by an MCPTT client are reduced and capped at 150 ms, and thus the average access time is reduced. As we explained in Section 3.2.1, when *TFG*1 expires, UE A assumes that the call of interest does not exist and will establish a new group call by itself, which marks the end of the MCPTT access time duration. However, given that the scenario simulated is B or C, i.e., group call *nn* exists already, these *TFG*1 expirations are premature, and the capped maximum access time might not be an improvement. Indeed, the MCPTT client, UE A, makes the wrong decision when establishing a new group call instead of joining the existing call. Consequently, the MCPTT client is not aware of the existing call, and multiple calls of the same group ID coexist. This undesirable situation of multiple calls with the same group ID will be referred to as the "multiple calls situation" for the remainder of this paper. The multiple calls situation will persist until the call merging procedure is triggered. After taking into account the extra delay introduced by the necessary call merging procedure, the average effective call control access time is very close to the results that we obtained when *TFG*1 is large, as seen in Figure 11, so the analytical model gives a good prediction of the effective call control access time after the call merge procedure. Simulation results with a sidelink period longer than 120 ms are not included in Figure 11 because UE A always makes the wrong decision in that case and, thus, the call merging procedure is always triggered to fix the multiple calls situation.

The disadvantages of the multiple calls situation include not only the extra delay due to the call merging procedure, but also the potential false impression/feedback to the MCPTT client. The MCPTT user, who is very likely a first responder on an urgent task, might think that no teammate is within communication range, although some teammates may be deployed nearby. Similarly, the MCPTT user may try to switch groups if he/she is notified wrongly that the group does not have an existing call. As another example, a first responder may decide to take actions at a higher risk, instead of waiting for support/backup from other teammates when in a time-sensitive or life-threatening situation. Therefore, although a smaller *TFG*1 configuration may reduce the wait before establishing a new call if the call does not exist, the 3GPP default setting of *TFG*1 = 150 ms is worth further discussion to reduce the occurrence of the multiple calls situation, and it may be desirable to take T_{SL} into account when configuring *TFG*1.

The negative impact of the multiple calls situation becomes more pronounced when the channel conditions degrade. Figure 12 compares the impact of channel loss on the call control access time when timer *TFG*1 is configured with the 3GPP default setting. When the channel PER of the fourth transmission attempt, $p4$, in the Sidelink Shared Channel increases from 0 (lossless) to 0.1 (lossy), $T_{CallControl}$ remains similar. This is because the potential long delay due to message loss is mostly absorbed by the expiration of *TFG*1. However, the value of $T_{CallControl}$ discussed above is only the nominal access time and does not consider that the MCPTT client might make the wrong decision to establish a new group call when *TFG*1 expires. When comparing the effective access time experienced by the MCPTT client, i.e., considering the extra delay introduced by the call merging procedure wherever applicable, the average effective $T_{CallControl}$ value has increased considerably due to the channel loss. The extra delay observed, which is quite large, is mainly due to the slow periodicity of the Call Announcement transmission cycle when the MCPTT UE is not responding to a Call Probe. When a Call Probe message is lost, or the responding Call Announcement message is lost, a UE is more likely to wait for the periodic Call Announcement to trigger a call merging procedure, which is slower.

Table 5 shows the probability of the "multiple calls situation" occurrence if *TFG*1 is set to 150 ms. Because of the potential packet loss due to the unnecessary retransmission of MCPTT messages, an error situation is more likely to occur

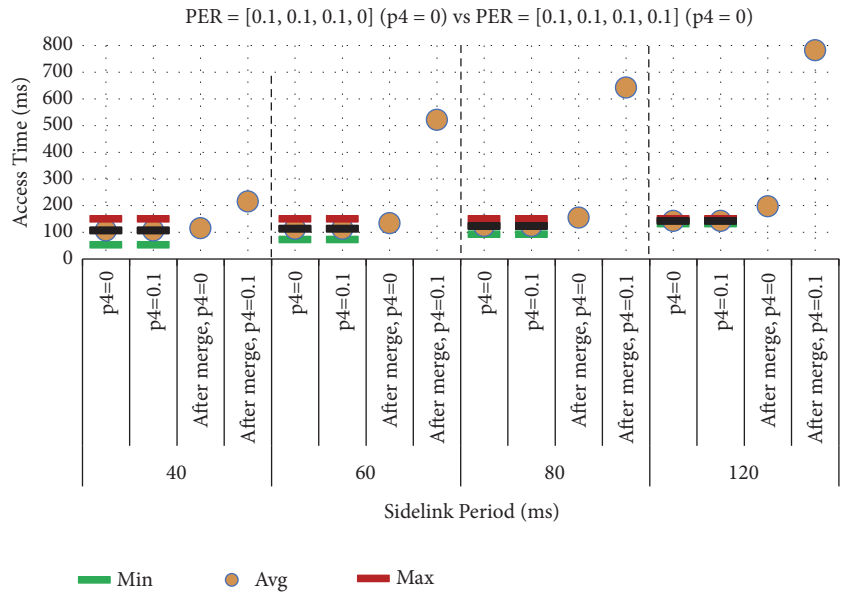


FIGURE 12: Impact of channel loss on call control access time with TFG1 = 150 ms with 95% Confidence Intervals for the average values from the simulations.

TABLE 5: Probability of "Multiple Calls Situation" for various $TFG1$ configurations and channel conditions ($TFG1$ = 150 ms).

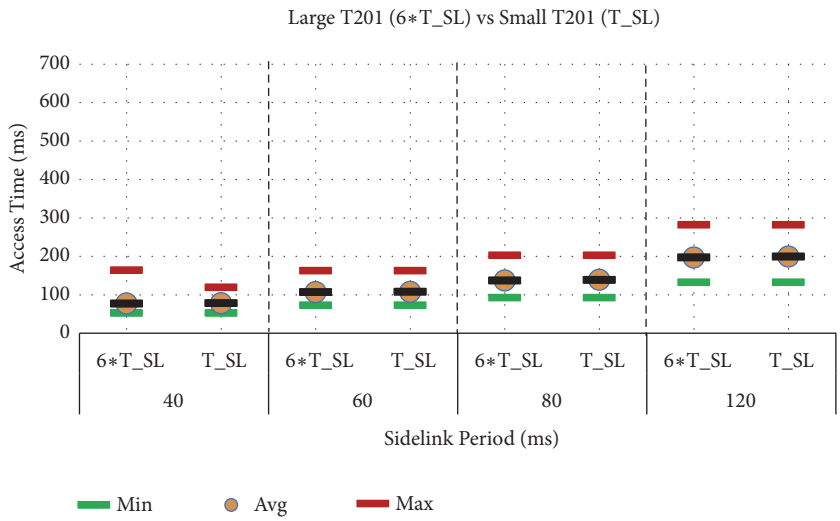
		T_{SL}(ms)	40	60	80	120
Probability of the occurrence of "multiple calls situation"	$PER = [0.1, 0.1, 0.1, 0]$	$TFG3 = 6 * T_{SL}$	0.169	0.384	0.504	0.894
		$TFG3 = T_{SL}$	0.176	0.378	0.513	0.904
	$PER = [0.1, 0.1, 0.1, 0.1]$	$TFG3 = T_{SL}$	0.186	0.362	0.525	0.917

when $TFG3$ is set to T_{SL} than when it is set to $6 * T_{SL}$. When the channel condition is worse, the occurrence rate of the "multiple calls situation" also increases, as shown in the last two lines of Table 5.

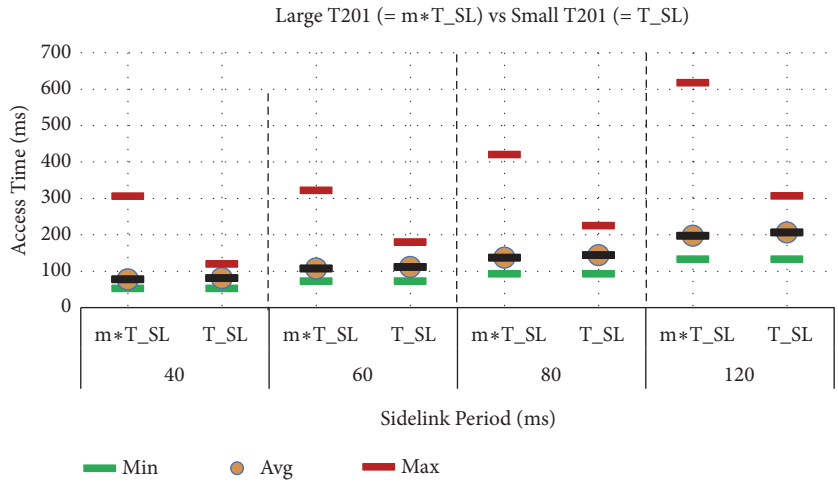
4.3. Impact of Floor Request Retransmission T201. The impacts of a smaller value for the Floor Request Retransmission Timer, $T201$, such as the 3GPP default setting T_{SL}, are multifold. On the positive side, if there is no floor arbitrator present, a smaller value for $T201$ enables the MCPTT user to declare the floor sooner, with the access time equal to $T201 * C201$, than when $T201$ is larger (e.g., $6 * T_{SL}$) as shown in Figure 13(a), or when $T201$'s value is based on the provision of the near worst case as shown in Figure 13(b). In the case shown in Figure 13(b), $T201 = m * T_{SL}$, where m is the smallest integer which is greater than $\max\{T_{FloorControl}\ (F2.1)\}/T_{SL}$, which can be calculated based on (18). By comparing Figures 13(a) and 13(b), we can see that the channel loss makes the positive impact of a small $T201$ configuration more pronounced. Note that we obtained the results in Figure 13(a) by running the simulation with the same assumption as for Figure 7, except that we configured $T201$ at the 3GPP default setting T_{SL} instead of at $6 * T_{SL}$. We obtained the results of Figure 13(b) by running the simulation with the same assumption as for Figure 7, except that we set PER to [0.1, 0.1, 0.1, 0.1] instead of [0.1, 0.1, 0.1, 0], and we configured $T201$ at either T_{SL} or $m * T_{SL}$.

However, on the negative side, the decision made by the MCPTT client (e.g., assume the floor is idle) with a smaller $T201$ value might be premature, and it may lead to an undesirable situation in which there are multiple floor arbitrators, which we will refer to as the "multiple floor arbitrators situation". Note that this situation is neither easy nor quick to resolve, given the current procedures. Table 6 shows the increase in ratio of "wrong decision" vs "correct decision" when comparing small $T201$ settings with large $T201$ settings. It is also clear from comparing Table 6(a) to Table 6(b) that the channel loss makes the negative impact of a small $T201$ configuration even more evident. Note that the results in Table 6 show the worst case outcome because we assumed that the incumbent floor arbitrator UE B does not generate RTP packets during UE A's access time. If the current floor arbitrator UE B is actively sending RTP packets, the counter C201 at UE A is reset every time an RTP packet is received. Therefore, it will take more time and will be less likely for UE A's $C201$ to reach the threshold and make a wrong decision when UE B is talking.

We encounter the "multiple floor arbitrators situation" when there is more than one floor arbitrator simultaneously in the same off-network group call, although the correct behavior is that there should be at most one floor arbitrator



(a) Access time when $T201 = 6 * T_{SL}$, PER = [0.1, 0.1, 0.1, 0] with 95% Confidence Intervals for the average values from the simulations



(b) Access time when $T201 = m * T_{SL}$, PER = [0.1, 0.1, 0.1, 0.1] with 95% Confidence Intervals for the average values from the simulations

Figure 13: Impact of $T201$ on floor control access time.

Table 6: Impact of $T201$ configuration—potential multiple floor arbitrators situation.

(a) PER = [0.1, 0.1, 0.1, 0]

	T_{SL} (ms)	40	60	80	120
Ratio of "premature $T201$ expiration case" (i.e., the wrong decision case) vs "the correct decision case"	$T201 = T_{SL}$	0.026	0.018	0.0007	0.0004
	$T201 = 6 * T_{SL}$	0	0	0	0

(b) PER = [0.1, 0.1, 0.1, 0.1]

	T_{SL} (ms)	40	60	80	120
Ratio of "premature $T201$ expiration case" (i.e., the wrong decision case) vs "the correct decision case"	$T201 = T_{SL}$	0.126	0.145	0.075	0.060
	$T201 = m * T_{SL}$	0	0	0	0

at any time. There are several circumstances that may lead to the "multiple floor arbitrators situation," such as a premature $T201$ timer expiration or the hidden node problem. The potential disadvantages of the multiple floor arbitrators situation are as follows:

(i) The lack of resolution procedure for "multiple floor arbitrators situation." This is in contrast to the "multiple group calls situation," for which the call merging procedure is already defined in the call control protocol as a remedy.

Table 7: Impact of $T201$ Configuration—potential half duplex effect.

	T_{SL} (ms)	40	60	80	120	160	240	280	320
Ratio of double occupancy vs single occupancy	$T201 = T_{SL}$	0.603	0.723	0.581	0.575	0.573	0.523	0.529	0.534
	$T201 = 6 * T_{SL}$	0	0	0	0	0	0	0	0
Ratio of samples discarded	$T201 = T_{SL}$	0.112	0.126	0.071	0.058	0.051	0.024	0.025	0.025
	$T201 = 6 * T_{SL}$	0	0	0	0	0	0	0	0

Table 8: Average MCPTT access time in milliseconds with 3GPP default settings.

T_{SL} (ms)	40	60	80	120	160	240	280	320
Scenario A	150	150	150	150	150	150	150	150
Scenario B	184.46	221.95	263.45	342.81	Call Merge procedure is always triggered			
Scenario C	226.43	293.82	364.47	503.17	Call Merge procedure is always triggered			
Scenario D	78.03	108.13	138.98	199.63	260.22	379.80	440.32	500.85
Scenario E	120	180	240	360	480	720	840	960

(ii) The MCPTT client who takes the floor prematurely is not aware of the situation and might start talking while not knowing that no one is listening;

(iii) The incumbent floor arbitrator is not aware of the situation either and does not know that one or more group members are not listening.

As was the case in the call control study, the negative impact of the Floor Request Retransmission Timer $T201$ being configured too small can be noticed in Table 7 and is mainly due to the unnecessary retransmissions of the Floor Request message.

When comparing the same performance metrics of the $TFG1 = T_{SL}$ case in Table 4 with the $T201 = T_{SL}$ case in Table 7, the call control procedure seems to perform slightly better than its floor control counterpart. This is because the $TFG2_{probe}$ timer adds a random delay to the transmission timing of the responding Call Announcement message. Consequently, the transmission of the Call Announcement message does not always compete with the retransmission of a Call Probe message for resources in the same sidelink period.

4.4. MCPTT Access Time with 3GPP Default Settings. Table 8 summarizes the simulation results of the average MCPTT access time in milliseconds for different scenarios when the MCPTT parameters are configured according to the 3GPP default settings.

We obtained the values in the table assuming no hardware or software processing time. In addition, we obtained the access times when UE A makes the correct decision as described in Table 3. However, wrong decisions are not uncommon occurrences with the 3GPP default settings, which might lead to the "multiple floor arbitrators situation" and/or the situation of multiple calls with the same group ID.

5. Conclusions and Future Work

We developed the analytical models and simulation tools described in this paper to estimate and evaluate the performance and user experience of MCPTT off-network mode operations over ProSe in LTE, with a focus on access time when an MCPTT UE is out of coverage. We defined performance metrics and application scenarios for the access time analysis. Our analytical models provide good estimates of the minimum, average, and maximum for both call control and floor control access times experienced by an MCPTT client under various scenarios, as shown by our simulation results. We summarize our main observations with respect to parameter configurations in Table 9. These observations may be used to understand the tradeoff and provide configuration guidelines for MCPTT and ProSe.

The sensitivity analysis results show that preventive solutions, such as proper parameters configuration and/or modified procedures, need to be investigated to reduce the possibility of an MCPTT UE's making wrong decisions. In addition, 3GPP may re-evaluate the default setting of timers TFG1, TFG3, and T201 due to the undesirable occurrence of the 'multiple floor arbitrators situation,' the coexistence of multiple group calls with the same group ID, and the relatively high probability of the half duplex effect.

Our future work includes enhancing the access time analysis to consider more aspects related to ProSe, such as a probability model of the half duplex effect in the PSCCH. In addition, we will evaluate the MCPTT access time performance for the scenarios when multiple active UEs try to talk at nearly the same time, or there are multiple ongoing group calls competing for radio resources over ProSe.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Table 9: Impact of parameters configuration on MCPTT access time.

Parameter	Small	Big
T_{SL}	(i) Shorter access time; (ii) More overhead;	(i) Longer access time; (ii) Less overhead;
$TFG3$	(i) Faster retransmissions of Call Probe; (ii) More unnecessary retransmissions of Call Probe (e.g., with 3GPP default configuration) are likely;	(i) Reduced unnecessary retransmissions of Call Probe; (ii) Necessary Call Probe retransmission may be slowed down;
$TFG1$	(i) Reduced nominal maximum and average access time; (ii) The call merge procedure is more likely to be triggered (e.g., with 3GPP default configuration), and result in increased effective maximum and average access time;	(i) Nominal maximum and average access time may be larger; (ii) Reduced unnecessary call merge procedure triggering, and thus reduced effective maximum and average access time;
$T201$& $(C201 * T201)$	(i) Faster determination of idle floor; (ii) The "multiple floor arbitrators situation" is more likely to occur (e.g., with 3GPP default configuration); (iii) More unnecessary retransmissions of Floor Request (e.g., with 3GPP default configuration) are likely.	(i) Slower determination of idle floor; (ii) The "multiple floor arbitrators situation" can be reduced ($C201 * T201$); (iii) Reduced unnecessary retransmissions of Floor Request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] National Public Safety Telecommunications Council (NPSTC), "Mission Critical Voice Communications Requirements for Public Safety," 2011.

[2] National Public Safety Council Telecommunications, "Public Safety Broadband Push-to-Talk over Long Term Evolution Requirements," Tech. Rep., 2013.

[3] A. Rebeiro-Hargrave, *OMA Push-to-Talk Architecture, Multimedia Group Communication*, Chichester, UK, pp. 23–57, 2AD, 2008.

[4] D. Ma, W. Huang, Q. Sun, W. Li, C. Yin, and Z. Shao, "Design and realization of the poc system with short session-setup-delay," in *Proceedings of the 2012 International Conference on Computer Science and Service System, CSSS 2012*, pp. 1075–1078, August 2012.

[5] E. Casini, N. Suri, M. Breedy, P. Budulas, J. Kovach, and R. Roy, "PeerTalk: A push-to-talk and instant messaging service for tactical networks," in *Proceedings of the 2013 IEEE Military Communications Conference, MILCOM 2013*, pp. 1476–1481, November 2013.

[6] M.-H. Tsai and Y.-B. Lin, "Talk burst control for push-to-talk over cellular," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2612–2618, 2008.

[7] A. Kuwadekar and K. Al-Begain, "An evaluation of push to talk service over ims and lte for public safety systems," in *Proceedings of the 2014 6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*, pp. 412–416, November 2014.

[8] Open Mobile Alliance (OMA), *Enabler Release Definition for Push to Communicate for Public Safety (PCPS)*, 2017.

[9] 3GPP, "Technical Specification Group Services and System Aspects; Mission Critical Push To Talk (MCPTT) over LTE; Stage 1 (Release 14); TS 22.179," Tech. Rep., 2016.

[10] J. Kim, O. Jo, and S. W. Choi, "State-based uplink-scheduling scheme for reducing control plane latency of MCPTT services," *IEEE Systems Journal*, 2018.

[11] R. Solozabal, A. Sanchoyerto, E. Atxutegi, B. Blanco, J. O. Fajardo, and F. Liberal, "Exploitation of mobile edge computing in 5G distributed mission-critical push-to-talk service deployment," *IEEE Access*, vol. 6, pp. 37665–37675, 2018.

[12] 3GPP, "Technical Specification Group Core Network and Terminals; Proximity services (ProSe) User Equipment (UE) to ProSe Function Protocol Aspects; Stage 3 (Release 14); TS 24.334," Tech. Rep., 2018.

[13] J. Wang and R. Rouil, "Assessing coverage and throughput for D2D communication," in *Proceedings of the 2018 IEEE International Conference on Communications, ICC 2018*, May 2018.

[14] D. W. Griffith, F. J. Cintron, and R. A. Rouil, "Physical sidelink control channel (PSCCH) in mode 2: performance analysis," in *Proceedings of the 2017 IEEE International Conference on Communications, ICC 2017*, May 2017.

[15] D. Griffith, F. Cintrón, A. Galazka, T. Hall, and R. Rouil, "Modeling and simulation analysis of the physical sidelink shared channel (PSSCH)," in *Proceedings of the 2018 IEEE International Conference on Communications, ICC 2018*, May 2018.

[16] A. M. Cipriano and D. Panaitopol, "Performance analysis of sidelink data communications in autonomous mode," in *Proceedings of the 2018 IEEE Wireless Communications and Networking Conference, WCNC 2018*, pp. 1–6, April 2018.

[17] National Institute of Standards and Technology, "Public Safety Communication modeling tools based on ns-3," 2018. https://github.com/usnistgov/psc-ns3.

[18] 3GPP, "Technical Specification Group Core Networks and Terminals; Mission Critical Push To Talk (MCPTT) call control; Protocol specification (Release 14); TS 24.379," Tech. Rep., 2018.

[19] 3GPP, "Mission Critical Push To Talk (MCPTT) media plane control; Protocol specification (Release 14); TS 24.380," Tech. Rep., 2018.

[20] P. Varin, Y. Sun, and W. Garey, "NISTIR 8236 Test Scenarios for Mission Critical Push-To-Talk (MCPTT) Off-Network Mode Protocols Implementation," Tech. Rep., 2018.

[21] 3GPP, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 14); TS 36.213," Tech. Rep., 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6206854, 13 pages
https://doi.org/10.1155/2019/6206854

Research Article

Processing and Communication Delays in EWS: On the Performance of the Earthcloud Prototype

Martin Klapez,[1,2] **Carlo Augusto Grazia**,[1,2] **Maurizio Casoni**,[1] **Simone Zennaro**,[2] **and Matteo Cozzani**[2]

[1]*Department of Engineering Enzo Ferrari, University of Modena and Reggio Emilia, via Pietro Vivarelli, 10, 41125 Modena, Italy*
[2]*Earthcloud Association, Italy*

Correspondence should be addressed to Martin Klapez; martin.klapez@unimore.it

Received 19 November 2018; Accepted 5 March 2019; Published 26 March 2019

Academic Editor: Mauro Femminella

A Seismic Alert System (SAS), also called Earthquake Warning System (EWS) or Earthquake Early Warning System (EEW or EEWS), represents one of the most important measures that can be taken to prevent and minimize earthquake damage. These systems are mainly used to detect P-waves and the faster seismic waves and to subsequently trigger an alarm about the incoming S-waves, the slower and most dangerous seismic waves. In some cases, distributed systems are also able to alert some locations before the impending P-waves strike them. This paper presents Earthcloud, a cloud-based SAS that aims to provide all the former capabilities while retaining financial accessibility. Earthcloud first results, generated from four months of data acquisition, are compared with those coming from other systems. In particular, the paper focuses on processing and communication delays, showing how the Earthcloud new detection strategy may minimize delays. Although a thorough test campaign with more sensor nodes is needed to assess performance reliably, especially for highly dense urban scenarios, initial results are promising, with total latencies for Earthcloud always kept under the 1-second mark, despite being at the expense of solid magnitude estimation.

1. Introduction

Due to the relatively fast propagation of earthquake waves, it is only possible to receive alerts from SASs in a timeframe that goes from a few seconds to tens of seconds before the strike. Nevertheless, they can potentially yield very significant benefits. If an earthquake occurs at night when people are usually sleeping, for instance, the primary risk for human health (primary in terms of the number of people affected) is represented by the possible collapse of buildings or parts of them. Potential victims usually wake up abruptly, most likely in a state of panic and with the risk of being overwhelmed by events. Risks for human health would be even higher if an earthquake occurs during the day when many people are working or traveling. Examples of these hazards, unfortunately, abound and can go from small injuries to large-scale nuclear disasters, for instance, derailment of trains on railroads and subways, vehicles on bridges and tunnels, dangerous machines and chemicals in work environments, suspended loads and work at heights in building sites, fires, and more, plus the collapse of buildings, as mentioned earlier. In addition to all of these, some risks involve machinery that may or may not affect human health but that would inevitably cause economic damage. Receiving a seismic alert, even if only seconds before the arrival of destructive waves, can prevent human losses, injuries, and damage to machinery and infrastructures. For instance, it can trigger the enforcement of a previously approved emergency plan that people can consist in moving away from hazardous equipment, taking cover under desks or load-bearing structures, etc. Machinery, on the other hand, can be automatically slowed, shut down, and/or isolated, to prevent damage as well as to avoid the ignition of cascading threats, such as fires that can kindle in the aftermath of an earthquake.

Designing and operating an effective SAS, however, is often not trivial, because an earthquake has to be detected, processed, and notified significantly faster than the propagation speed of its seismic waves. Furthermore, environmental,

traffic, industrial, and other kinds of noise that may easily generate false positives are often present and have to be effectively classified in a very short time with a very low error rate. This paper presents the second iteration of Earthcloud, an Internet of Things (IoT) SAS designed to be low-cost, low-power, and cloud-based. From the detection of the first anomaly to an alarm, processing and communication delays of several systems is here analyzed and compared with the results obtained from the Earthcloud prototype.

To provide the reader with solid interdisciplinary bases, Section 2 pinpoints related work, while Section 3 summarizes the necessary geological concepts. The deployed prototype is described in Sections 4 and 5, where sensor units and the cloud system are detailed, respectively. Section 7 presents the outcomes of four months of processing, with an emphasis on processing and communication delays. Section 8 concludes the writing.

2. Related Work: An Overview on SASs

The goal of this Section is not to be exhaustive; it is instead to present an indicative group of works from academia and industry, to form a representative overview of the options and to point the reader to relevant sources.

Back in 1996, Leach et al. wrote one of the first works on SASs based on computer technology [1]. Authors designed a neural network and trained it with real data extracted from seismograms. Then, the neural network has been fed with other data from the same sources, and its outcomes have been studied. An actual system was not developed, but the algorithm was successful in the accurate estimation of hazard start, stop, and duration time. The algorithm was able to generate a warning signal 0.3 seconds after detecting the first ground motion.

The majority of SAS prototypes reported in the literature aim to be low-cost use MEMS (i.e., Microelectromechanical System) accelerometers as sensors. In [2], for instance, Hoque et al. built a prototype where sensors are composed of accelerometers attached to Arduino microcontrollers. The control system is a LabVIEW software on a laptop that communicates with sensors through the ZigBee standard. A ZigBee antenna can theoretically have a range up to hundreds of meters, but, generally, actual range is in the order of tens of meters. The latter applies to the system, as presented by the authors. With respect to Earthcloud, it is also moderately more complex to set up. From the paper, it is not clear whether the prototype has been deployed or not. Another work that leverages accelerometers and Arduino microcontrollers is that presented by Sherki et al. [3]. Instead of ZigBee, a GSM module is employed, and the system is tested against a vibration generator machine. The focus of the authors, however, is in this case on signal analysis.

Accelerometers embedded in smartphone have also been extensively used. For instance, Uga et al. [4] test these sensors specifically for SASs applications and study how to separate false positives from real ground motion, although tests are performed against artificial data obtained with a shakeboard, and smartphones are positioned on horizontal flat surfaces for the entire duration of the experiments. Shakeboards are also employed to test prototypes designed to form an IEEE 802.11 (WiFi) Wireless Mesh Network [5, 6]. In these cases, however, authors focus specifically on the performance of wireless communications. They found that the sudden small-amplitude P-waves (see Section 3) shaking can have a huge impact on their performance, especially with no or scarce Line Of Sight between transceivers. Accelerometers in smartphones have also been coupled with cloud elaboration before. Heryana et al. [7] rely on 2G/3G/4G cellular technologies or IEEE 802.11 to pass data between devices and a cloud service. However, most of the paper is spent detailing the software development cycle of the proposed Android application, and very few information can be found about the effectiveness of the system.

Other works focus on unconventional systems. Shayo et al. [8] use the same low-cost accelerometer sensors seen in other solutions. However, sensors transmit data to a computer that, in the event of an emergency, transforms the alarm signal in SMS form and then send it through cellular networks. Another workstation has been set up to receive those SMS and measure the elapsed time. Authors found that, although the delays varied, on average these exceeded the maximum tolerable delay for a SAS and concluded that SMS is not a reliable platform for those means. An unorthodox sensor is presented by Heindl [9]. Author proposes to detect ground motions through the variations in read/write error rates of common electromechanical magnetic hard disks. The distributed system is designed in a P2P fashion. As in [7], the SAS would require substantial user collaboration to work correctly. Alternatively, although vendors would probably oppose such a choice for security and privacy reasons, operating systems might be modified to perform those SAS tasks in the background, without the users' consent. Anyway, the system may become weak in the long term, as magnetic hard drives are replaced by faster solid-state storage devices.

About systems that rely on users collaboration through software deployed on common devices, the Earthquake Network Project by Finazzi [10] is one of the best examples, claiming 4 million downloads of its Android application, 750000 active users, and more than a 1000 early warnings sent. While the analysis of data coming from the network of smartphones is not trivial, statistical approaches are being studied to improve the system [11].

An example of SASs from industry is ShakeAlarm by Zaicenco et al. [12], developed by Weir-Jones Engineering Consultants, a system that detects P-waves (see Section 3) and claims to be able to determine in less than 0.5 seconds if following waves will be dangerous. ShakeAlarm is deployed in some regions of Canada and of the USA. There are also SASs developed by institutions and deployed on a large scale. In Japan, the Japan Meteorological Agency operates an EWS [13] mainly formed by about 300 single-function and multipurpose seismometers; the latter are equipped with satellite mobile phone communication capability for backup purposes and a power supply that can keep the whole system operational for about 72 hours in the event of power failure. The Japanese EWS fetches data also from seismometers managed by universities, by the National Research Institute

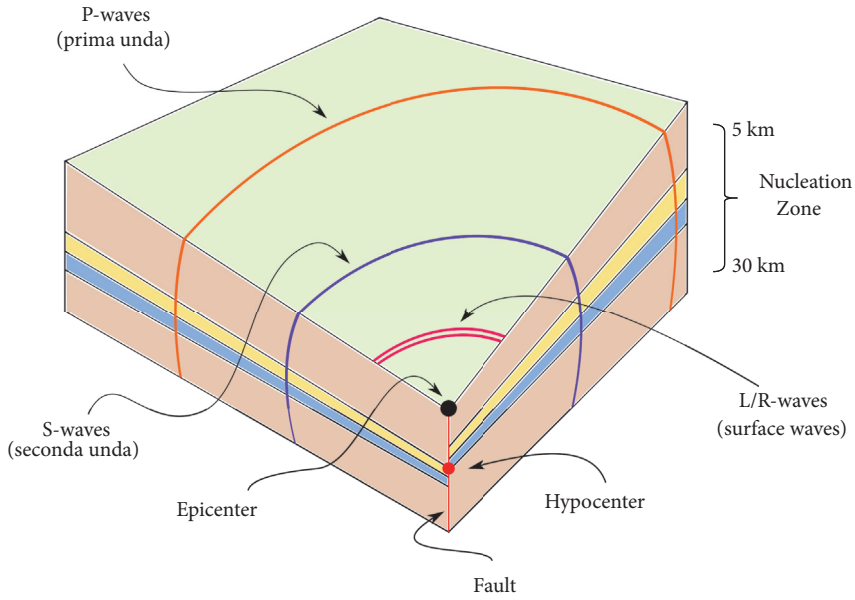


Figure 1: Basic pattern of an earthquake.

for Earth Science and Disaster Resilience, and by the Japan Agency for Marine-Earth Science and Technology. Taiwan also has a nationwide SAS [14], while Mexico has a partial EWS called SASMEX that comprises the SAS of Mexico City, in continuous operation since 1991, and the SAS of Oaxaca City that started its services in 2003. In the United States, the west coast is currently covered by an experimental system called ShakeAlert, presented by Burkett et al. and Given et al. in [15], [16], respectively.

Earthcloud is a low-cost, low-power, and cloud-based SAS at the prototypal stage. Klapez et al. previously presented a version [17] that had two processing layers. The first was represented by sensor nodes and the second by the cloud infrastructure. New data was continuously generated by the sensor devices and continuously sent to the cloud system. The same data was processed twice, once by the sensors, once by the cloud. Both were able to issue an alarm, on the basis of the respective processing results. As the processing power in the first layer is limited, and to keep processing times acceptable, the cloud infrastructure was responsible for issuing the most reliable warnings. In this paper, we present the second iteration of Earthcloud. As it is described in Section 5, it embeds a fundamental change in how warnings are triggered and on how data is handled.

3. Key Geological Concepts

Earthquakes are generated in an area called the nucleation zone. Most of the times, this area is located inside the seismogenic layer, that is the part of the Earth's crust with a depth ranging from 5 to 30 km. In the nucleation zone, multiple types of waves are generated, with different characteristics based on destructive potential and speed.

These waves mainly come in two arrivals. The time between the two is the opportunity to detect and confirm the earthquake and send an alarm signal. Each of these two arrivals consists of many individual elastic waves that have traveled from the epicenter to recombine at the recording site as a function of their respective velocities, focal distances, and propagation paths. As depicted in Figure 1, waves belong to two types: body waves and surface waves.

The first arrival is composed by the fastest of body waves, the P-waves (from Latin *prima unda*, i.e., primary waves). These are compressional or longitudinal and shake the ground in the direction of their propagation using compression or rarefaction, while their speed is between 4 and 8 km/sec. P-waves are usually not destructive.

Body waves and surface waves may compose the second arrival. They often produce both horizontal and vertical ground motion and their peak velocities, peak accelerations, and duration in time may cause significant damage to structures.

The body waves in the second arrival are called S-waves (from Latin *secunda unda*, i.e., secondary waves); they shake the ground in a direction that is perpendicular to the direction of propagation, while their speed is about 60% to that of the respective P-waves [2]. S-waves can be destructive in up to several kilometers from the epicenter due to the phenomenon of seismic amplification, linked to the local geology and morphology.

Surface waves are categorized into two types, called Love and Rayleigh, from the name of the scientists that modeled them. They are formed by constructive interference between P-waves and S-waves, and they are the most dangerous. Love waves travel slower than both P-waves and S-waves; Rayleigh waves are the slowest. However, while they are slower, surface waves decay much less with the distance than

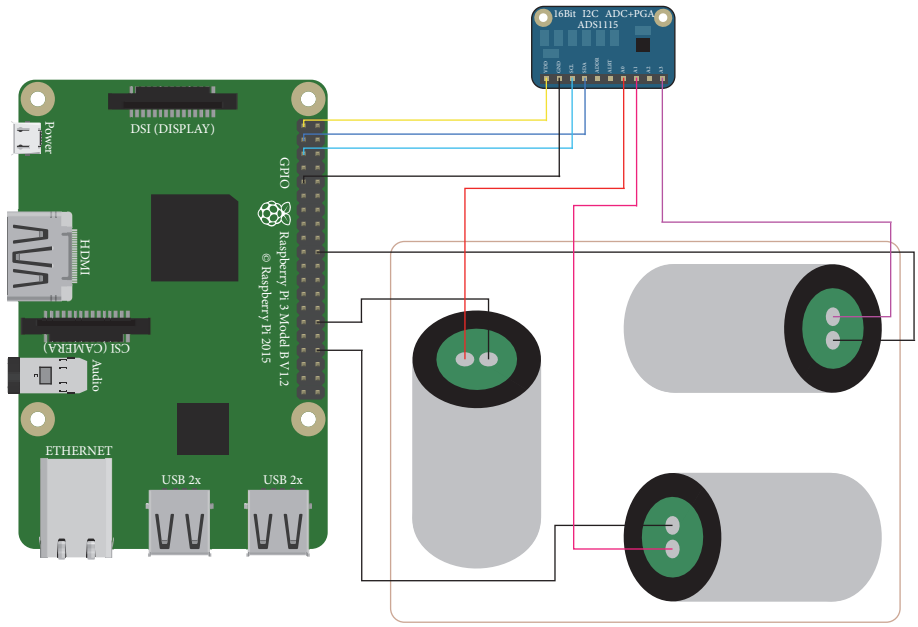


Figure 2: An Earthcloud Sensor Subsystem.

body waves, as they mainly travel on one axis instead of three. The Earth's crust where they travel acts as a waveguide that provides little attenuation loss. In big earthquakes, they may circumnavigate Earth several times before being completely dissipated. In general, surface waves cause more movement than body waves. In addition, the interaction among them as they propagate can produce considerable amplification of ground motion near the surface, a phenomenon called the free surface effect that occurs when upgoing and downgoing reflected waves are in phase and of considerably greater wavelength than the thickness of the crust.

4. Earthcloud: Sensor Devices

An Earthcloud sensor system is composed of three elements: a Raspberry Pi 3 Model B V1.2, the Adafruit ADS1115 Analog-Digital Converter (ADC), and a set of three 4.5 Hz geophones. Figure 2 depicts and details elements and the wired connections among them. Size proportions are real. Raspberry Pis mount a Linux distribution as the operating system; specifically, they mount Raspbian Stretch 9.4 with Linux kernel version 4.14-34 (4.14.34-v7+).

As a whole, a single sensor system needs less than 1W. To estimate the power consumption of the Raspberry Pi boards, we based our calculations on the specifications [18] published by the Raspberry Pi Foundation, in particular, those regarding power consumption of the same specific board in idle and low-load conditions. From an average current of ~ 300mA, we removed 50mA for the HDMI port and 100mA for mouse and keyboard (we picked up the best case, the aggregated figure can go up to 1A), all components not used in our setup. The values in [18] include WiFi usage; we assumed a similar consumption for the Ethernet port, that is the one used in our setup. All this, when considering that the board is powered with 5V, yields a worst-case power consumption of 0.75W. The GPIO pins may safely draw up to 50mA overall, while an individual GPIO pin can only safely draw 16mA. In our setup, only the cable pictured in yellow in Figure 2 draws a significant amount of current to power the ADC through the 3.3 VDC GPIO pin; the power needed for that is, therefore, equal to 0.0528W at most. The blue(ish) wires transmit data from the ADC to the Raspberry Pi board only, and, therefore, for the latter, those pins need a negligible amount of power. Geophones are instead passive devices that only need to be grounded. System-wide, this amounts to a power upper bound of ~ 0.8W; considering common power supply efficiency around 80%, we can safely claim a total power consumption value equal to or less than 1W.

The geophones are the Earthcloud sensor elements. They usually sit in between accelerometers and seismometers in function and price, the latter being the most accurate, delicate, and expensive system. A typical geophone has one working axis and has the function to convert ground movement that has the same direction of the working axis into voltage. Therefore, any voltage deviation from the baseline is data. A typical geophone consists of housing and, inside, of a mass suspended by means of mechanical springs. The spring-mass system has a resonance frequency, also called natural frequency, in the working axis of the geophone. Conceptually, when the sensor moves, if the velocity produces a frequency that is lower than the natural, both housing and mass move. If the generated frequency is higher, only the housing moves and the mass tends to hold its position.

In modern geophones, like those used by Earthcloud, the mass is formed by a cylindrical frame in which, externally, two coils are attached so as to surround the cylinder. The double coil allows reducing data distortion. The cylinder and coil assembly are usually supported by springs, directly

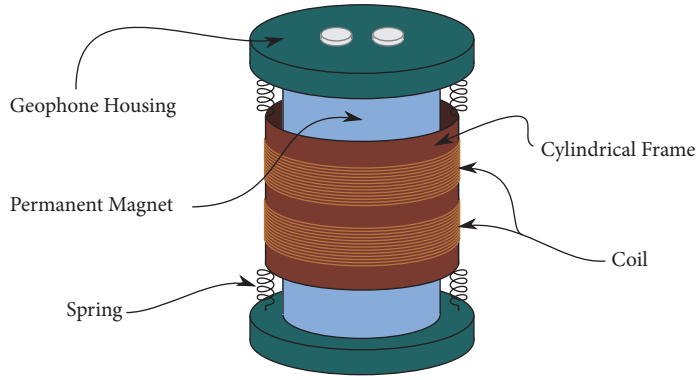


Figure 3: Inside an Earthcloud Geophone.

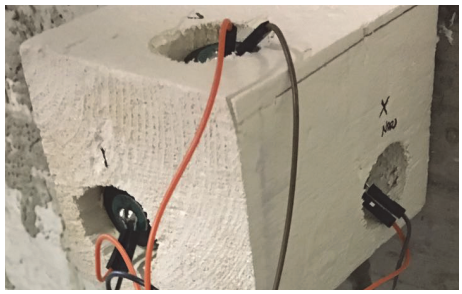


Figure 4: Geophone Housing.

connected to the geophone housing. Inside the cylinder frame, there is a permanent magnet also directly attached to the housing that does not move. Figure 3 illustrates a typical geophone of this kind. Figure 4 shows the geophones in an Earthcloud sensor system, in which they are enclosed in a rigid structure firmly mounted upon a surface like a wall.

Therefore, if the movement frequency is lower than the natural, there is conceptually no internal movement at all if considering the relationship between coils and magnet; i.e., the variation in the position of the coils with respect to that of the magnet is zero. If the frequency exceeds the natural, coils tend to maintain their position while the magnet moves with the housing; therefore, there is a shift in location between coils and magnet. As the coils are in the field of the magnet, this shift produces a variation in voltage that is proportional to the movement velocity. If the geophone is fixed to the ground, its response is (if above natural frequency) proportional to ground velocity. For comparison, MEMS accelerometers respond to ground acceleration.

Geophones also have what is called a spurious frequency, usually due to the small movements that the mechanical springs can have in directions perpendicular to the working axis of the sensor. These springs, in fact, are designed to move linearly in the working axis but also to have the possibility of small movements in the plane perpendicular to the working axis. Motion in this plane, either transverse or rotational, is essential to allow freedom of mobility in the working axis for the cylinder and coil assembly. As there is a resonance frequency in the working axis, called natural frequency, there is also a resonance frequency in the axis perpendicular to the former, that is called spurious frequency (actually, there are spurious frequen*cies*, the lowest is taken as the spurious). The springs are, of course, very stiff in the direction perpendicular to their working axis, hence the high-frequency nature of the spurious resonance [19]. As shown in Figure 5, the natural frequency represents the lower bound of the usable bandwidth, while the spurious frequency the upper bound. Therefore, the two frequencies together determine the useful frequency range of the sensor. For all frequencies in this bandwidth, the output sensitivity (i.e., the smallest variation in voltage that can be measured by the sensor) remains approximately constant.

As it can be seen from Figure 5, the actual response of a geophone is not an on-off Boolean function. The mass still moves below the resonance frequency, but output resolution (i.e., the smallest variation in space that can be measured by the sensor) drops fast (output resolution is calculated by dividing sensitivity over noise, S/N). Inverse filtering can compensate by flattening the response below the natural frequency of the geophone, but it is only useful if there is an adequate signal-to-noise ratio. On the other hand, for frequencies higher than the spurious one, springs introduce additional resonances that generate noise and therefore lower resolution. Inconsistencies in materials and manufacturing processes have a substantial effect on the determination of the spurious frequency, hence its common specification as $SF \geq x$, which indicates that the actual frequency band in which spurious resonance will occur may be x or a $y > x$.

Geophones usually maintain signal digitization as a separate process, as the analog output is generated in the sensor first and then is sent to an external digitizer. For comparison, MEMS accelerometers derive their feedback from within the digitizing process instead. In general, accelerometers have lower noise at high frequencies, while geophones have lower noise at low frequencies. The signal itself actually degrades slower with MEMS accelerometers, usually around a 6-dB as the frequency halves. For comparison, the same figure for standard geophones is about 12-dB [20]. This should make those accelerometers good candidates for low-frequency recording; if the signal is strong, this is indeed the case. The problem with MEMS accelerometers is the intrinsically

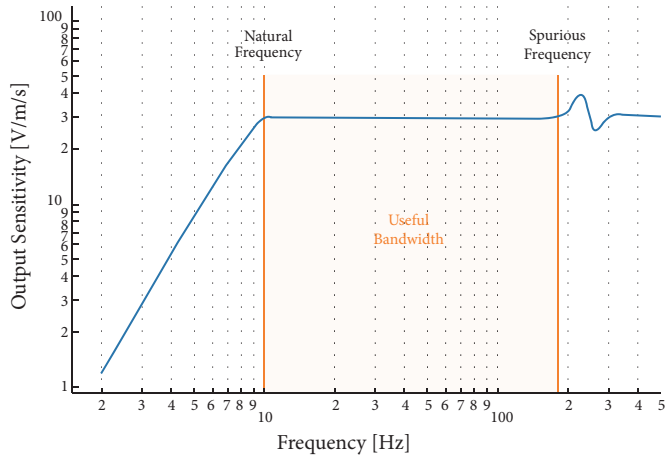


Figure 5: A transfer function for a typical 10 Hz geophone, with ~60% damping and 180 Hz spurious frequency.

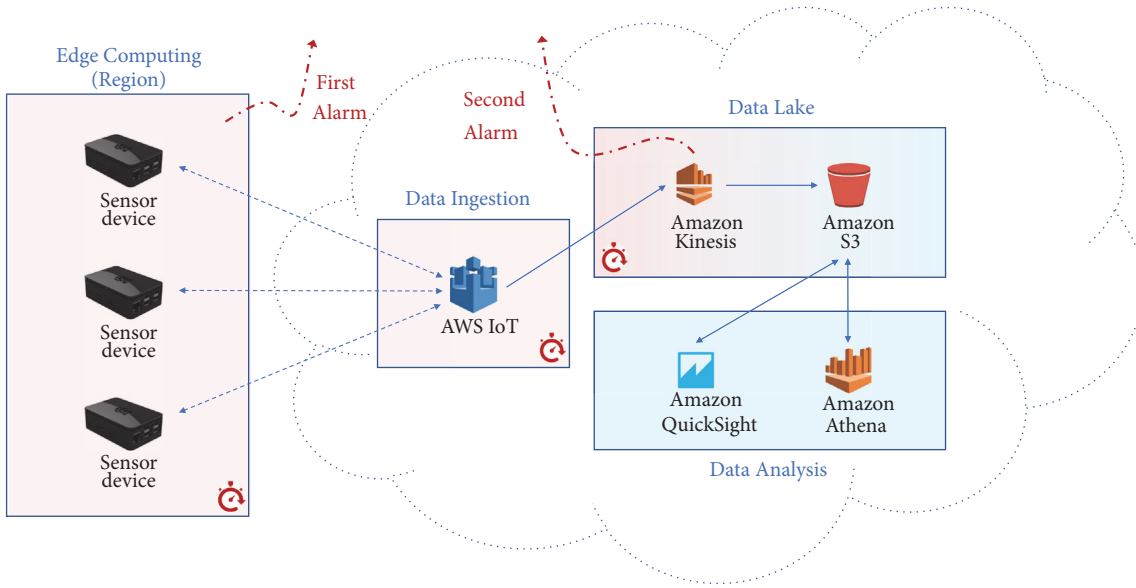


Figure 6: Earthcloud V2 Architecture.

high self-noise which makes them unsuitable for accurate low-frequency signals detection. Table 1 lists the technical specifications of the geophones currently used by Earthcloud. As shown in Figures 2 and 4, an Earthcloud sensor includes three geophones for three different axes that permit to gather data in a 3D domain.

5. Earthcloud: Cloud Infrastructure

Figure 6 shows the overall architecture of Earthcloud together with the composition of cloud services. At the moment, the latter are based on the Amazon AWS platform, in which Earthcloud services mainly operate from Ireland, EU.

IoT is a paradigm based on the connection between devices in the physical world, often sensors and data-gathering devices, and the Internet. Each Earthcloud sensor

Table 1: Technical specifications of the geophones used.

Name	EG-4.5-II
Natural frequency	4.5 ± 10% Hz
Damping	0.6 ± 5%
Coil resistance	375 ± 5% 1 Ω
Open circuit sensitivity (V/m/s)	28.8 ± 5%
Harmonic distortion	≤ 0.2%
Typical spurious Frequency	≥ 150 Hz
Moving mass	11.3 g
Max. coil excursion p-p	4 mm
Element diameter	25.4 mm
Element height	36 mm
Element weight	86 g
Operating Temperature	-40°C to +100°C

represents the edge of the cloud system, and it continuously scans the data incoming from its geophones. The only processing mechanism is a threshold-based filter; i.e., if data coming from a geophone exceeds a certain threshold (e.g., $v \geq 0.5mV$), a prewarning is issued to the cloud, together with the data that generated it. The principle on which Earthcloud warnings are based is purely probabilistic. In order to minimize to the maximum processing delays, sensors do not attempt to determine if signals from geophones are due to earthquake waves or noise. The system relies instead on the number of prewarnings received. As the number of sensors issuing a warning in the same region and in the same timeframe increases, the probability of contemporaneous false positives decreases.

Data generated by the Earthcloud sensors are encapsulated in MQTT messages continuously fed to AWS IoT, which preprocess and route them to Amazon Kinesis. When the data coming from its geophones exceed a set threshold, the sensor device sends a warning to AWS IoT, which in turn broadcasts it to the other sensor devices located in the same region. If a sensor device in a region issues a warning and receives a certain number of warnings with which it recognizes other n sensor devices of the same region as the source, it emits an alarm to registered entities. Subsequent alarms from other sensor devices located in the same region can be then safely ignored for a certain timeframe. Sensor data is small in size but can be continuously generated, from, potentially, thousands of sources. For Earthcloud, Kinesis has a dual function: to process data and issue alarms in real-time, and to act as a bridge between AWS IoT, the data entry point, and Amazon S3, the data endpoint. Clearly, the *real-time* part (depicted in Figure 6 with a light red background) is critical for Earthcloud. After receiving data for several warnings from the same region, Kinesis also issues an alarm to registered entities, serving as the confirmation that alarms are relevant and not false positives. Due to the significantly higher processing power of Kinesis with respect to sensor devices, in fact, the former can process more data and extrapolate more precise information, all in a short timeframe. The Kinesis alarm is supposed to follow those arriving at registered entities from sensor devices. However, there might be corner cases in which the latter fail to identify an earthquake; in these cases, the Kinesis alarm would be the first that is received by registered entities. Through an assessment of an ever-increasing data set, we expect to determine the probability with which Kinesis would issue a false alarm while being the only entity raising the issue. If we will find it to be sufficiently low, the alarms coming from the cloud system alone could be considered final even without the explicit support of sensor devices.

If needed, data can be converted into different formats before leaving the Kinesis bridge. Amazon S3 is a data storage service, well integrated with the other services from the same vendor. Data in S3 can be then further analyzed, manually and automatically. Amazon Athena allows querying S3 data through regular SQL (S3 data are stored as objects, not as SQL rows), while Amazon QuickSight, or other solutions, eases or performs data analysis. All the passages from Amazon S3 forward are regarded as batch processing (depicted in Figure 6 with a light blue background); more sophisticated analyses can be performed, but it can take anywhere from few seconds to few minutes to obtain the results. Although it is in the works, we do not have a system based on machine learning implemented at this stage. However, as the data set grows, all relevant data collectable from S3 becomes more valuable to train a suitable neural network adequately.

5.1. Network Performance Considerations. Cloud services are provided with already built-in redundancy features, at least as long as they are deployed on premises of established vendors. However, as strong earthquakes are likely to disrupt infrastructures [21–24], redundancy should be introduced to sensor devices too. As we used flexible, low-cost, and low-power devices, it should be feasible to power them with small rechargeable batteries as well as to enable on them multiple connection channels [25, 26], all while retaining financial convenience. Within shared wireless networks, it is important to maintain the latency low. To do so, the signal-to-noise ratio of the wireless network chain (i.e., sensor devices, access point, repeaters, routers, etc.) must be kept high, while the saturation of channels belonging to devices of the same chain must be avoided. To have high-throughput capabilities is not crucial per se for sensor devices, as the data rate they produce is very low. However, if a wireless channel starts to be congested, jitter and latency increase, even to the order of seconds. It is therefore important to either prevent [27, 28] or manage [29–33] congestion.

6. Earthcloud: The Prototype Setup

The first prototype of Earthcloud consists of the cloud system (in beta) and three devices. These have been deployed in Modena, an Italian city in the Emilia-Romagna region, a mid-north area of Italy, EU, classified as a moderate-to-high seismic zone according to the 1999 Global Seismic Hazard Map [34]. In 2012, a seismic swarm struck Modena's area with intense earthquakes [35]. The first strong earthquake was registered on May 20, 2012, at 02:03:52 UTC (04:03:52 local time, i.e., at night). It had a 5.9 Richter magnitude (M_L, for local magnitude) and a hypocenter located 6.3 km underground. Other subsequent earthquakes followed, two with 5.1 M_L. Seven people died, 50 were injured, and 5000 lost their houses, while many historical buildings collapsed, together with several farms and factories. Consequential to the disaster, soil liquefaction also caused collapses of recently built structures. On May 29, 2012, a second strong earthquake hit the same region with a 5.8 ML and a 10.2 km-deep hypocenter, at 07:00:03 UTC (09:00:03 local time, i.e., in the morning). Other subsequent earthquakes followed, among which one with 5.3 M_L and another with 5.1 M_L. Consistently with what is exposed in Section 1, damages were more significant. 20 people died, 350 were injured, and 10000 more lost their houses. Among the aftershocks that followed, on June 3, 2012, another one with 5.1 M_L hit the area. There were

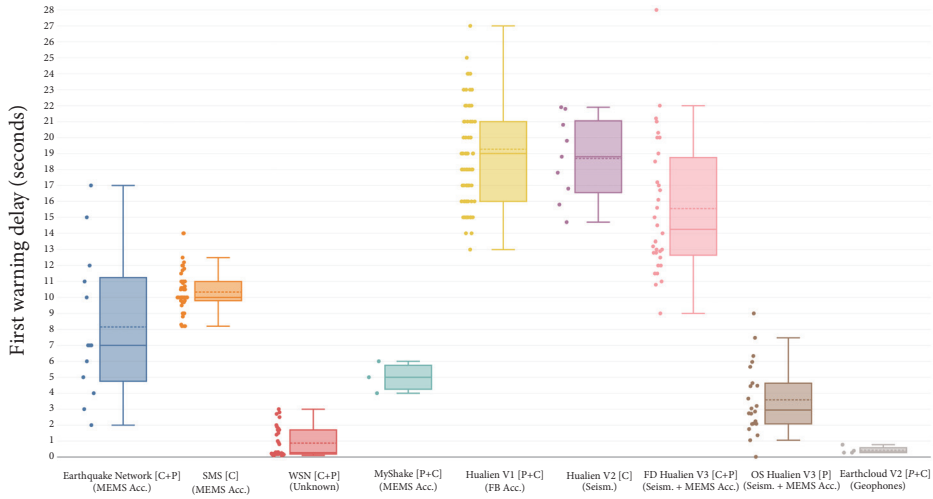


Figure 7: Comparison of first warning delays among different SASs (both prototypes and in production).

building damage and collapse, but no casualties. The two strongest earthquakes were also felt in nearby European countries, in particular in southeast France, Switzerland, south Germany, Austria, Slovenia, and Croatia. Damages were classified on the Mercalli scale (EMS-98) that classifies earthquake damages instead of its released energy, with a value of 8/12.

Two sensors have been initially installed on June 19, 2018, at the Department of Engineering Enzo Ferrari of the University of Modena and Reggio Emilia, inside a temperature-controlled server room. Devices are connected to the Internet through Ethernet cables, directly attached to the department switches for minimum latency, and are powered through electrical outlets. The geophone housings have been firmly mounted directly on walls concrete, with an industrial-grade wood-concrete glue. One has been installed on a load-bearing wall, while the other on a wall next to the room door, which is opened and closed very infrequently. The intent was to have two sensors in the same place but with one more sensitive to false positives. A week later, a third sensor has been deployed in a storage room belonging to a dismissed industrial building, located in a different area of the same city. The geophone housing was mounted on a load-bearing wall. This device was also connected to the Internet through an Ethernet cable, but with the difference of the latter being attached to a commercial off-the-shelf router manufactured for domestic purposes. Power was again provided by an electrical wall outlet. For the first prototype, unrefined ad hoc containers for geophones have been built by creating several rectangular prisms of approximately the same size from solid wood. Three holes to house the geophones have been drilled in each container, two for horizontal axes and one for the vertical axis, as in Figure 4. The holes allow for very limited movement of the geophones inside the housing, around 1 mm. As data analysis gets refined, more sophisticated housings will be created, in order to further minimize geophone allowances inside their containers or cancel them altogether.

7. Processing and Communication Delays

This section compares the latency performance of Earthcloud with the delays reported by different SASs in the literature. When possible, it is reported data from real earthquakes. Where this kind of data is not available and if simulation data is instead present, the latter is reported here.

Figure 7 depicts the comparison, where the vertical axis renders the delay that exists between the first wave arrival and the issue of the first warning, while the horizontal axis reports the SASs. In the label identifying a SAS, the round brackets in the second line inform about the kind of sensors employed by the system. The square brackets instead contain whether the delay values include processing delays (P) and communication delays (C). The order of these delays in the square brackets matters. The processing delay, in particular, can be found before or after the communication delay, depending on whether the processing performed on sensors is more significant or that on a remote system. In the case of Earthcloud, the symbol P is italic, to represent an almost negligible contribution. In the case of Hualien V3, the acronyms FD and OS stand for front-detection and on-site, respectively. A front-detection system aims to detect earthquakes in one place and give early warnings for more distant locations. On-site systems aim to recognize P-waves and to issue an alarm, in the same place where they are located, before S-waves arrive. The boxes of the box plot are composed as follows. The full horizontal line inside the perimeter of the boxes is the median, while the dotted line is the arithmetic mean. The lower and upper bounds of the boxes mark the first and the third quartiles, respectively. The whiskers include instead the values between $Q1 - (1.5 * IQR)$ and $Q3 + (1.5 * IQR)$, where IQR is the interquartile range, calculated with $IQR = Q3 - Q1$). We considered the other values, i.e., the dots outside the whiskers bounds, as outliers.

7.1. Earthquake Network. The Earthquake Network is a research project [10, 11] that leverages the MEMS accelerometers embedded in common smartphones to create a

crowdsourced front-detection EEWS. The principles on which the Earthquake Network operates are similar to those of Earthcloud. For the former, sensor nodes are represented by smartphones where a specific application is voluntarily installed by the user. If the smartphone is monitoring the output of its accelerometers, it sends an "alive" message to a central server every 30 minutes. In this way, it is possible to estimate how many sensors are active in each moment and where are they located. Similarly to Earthcloud, the processing in sensor nodes is kept to a minimum. In the Earthquake Network case, every vibration detected by a sensor appears to be logged, filtered, and then eventually sent to the central server together with the position of the smartphone. The sensor nodes, therefore, use the resources needed to operate their embedded accelerometers, log their data, filter it, and send it to the central server. The delay between the beginning of the ground shaking and the vibration detection from a smartphone is reportedly equal to $1.5s$ [11]. The detection algorithm is completely deployed server-side. In general, the authors report several figures about the delay experienced by various subnetworks during real earthquakes, including also false alarms. In particular, it is reported in detail the system response in the case of the M7.3 earthquake of May 12, 2015, in Nepal. The detection happened after approximately $6.5s$ from the moment the seismic waves reached the first smartphone. Firstly, the aforementioned $1.5s$ are assumed. As communication delay, $0.5s$ are assumed of having been needed for each one-way communication between smartphone and server, yielding a RTT (Round Trip Time) of $1s$. The server was able to confirm the earthquake after $4s$ from when it received the first notification. In general, the authors report total delay values in the range from 2 to 17 seconds, with a median of $7s$ and a mean value of $8.15s$. The paper [11] also contains delays resulting from simulations, not included here as per the guidelines defined in the incipit of Section 7.

7.2. SMS. Authors in [8] evaluate Short Message Service (SMS) messages as a platform to transmit seismic alerts. They created a prototype able to deliver SMSs through three mobile operators in Tanzania, finding that the communication delays fluctuate strongly, almost always exceeding the threshold of $5s$ took as reference. The delays resulting from the study have been deliberately isolated as communication delays only. Data is generated by a MEMS accelerometer connected to a sender computer and is then transmitted as SMSs to a receiver computer, every 20 minutes, for 21 days. It is worth to note that the same SMS was always simultaneously transmitted via the three operator networks. The study specifies that values greater than $15s$ have been discarded, as considered too high to be useful. The data reported here is composed by the values registered in one of the days of the experiment; specifically, the day with average values overall among the days of the trial. In this 24 hours, authors report total delay values in the range from 8 to 14 seconds, with a median of $10s$ and a mean value of $10.34s$.

7.3. WSN. In [6], Wireless Sensor Networks (WSNs) are proposed as low-cost tools to realize front-detection EEWSs. The authors refer to WSNs as "computer networks whose nodes communicate wirelessly using a license-free spectrum in a self-organized manner". This work focuses on the optimization of the network routing protocol, as it is determined that shakings representative of P-waves of M6 or greater at a distance of $40km$ or less can easily result in severe multipath and shadow fading effects that considerably affect performance of the WSN wireless communications right when the EEWS is needed the most. The paper mainly presents outcomes of simulations. Delays range from 0.1 to 3 seconds, growing together with the number of WSN nodes. As the latter increases, the reliability of the warnings decreases. The median results equal $0.3s$ while the mean $0.9s$. In the study, values "as low as $4s$" are cited for real implementations. The simulated values are very low. This is due to the fact that, like Earthcloud at its current stage, sensor nodes do not perform substantial processing. Unlike the Earthcloud prototype that can estimate the epicenter location, both the magnitude and the epicenter location estimations are not performed by WSN. There are some additional remarks that need to be made. Firstly, the reported figures are only correlated to the time interval between a generic P-wave arrival and a generic positive alarm decision, but information about which node does what is missing. Negative delays are also reported in the paper, clearly representing a warning arriving to a node still extraneous to the P-wave detection. This might strongly skew this system results in the framework of the comparison presented here. To mitigate this problem, only the positive values have been considered, although it had not been possible to correlate them correctly. On the other hand, the values cited to be "as low as $4s$" for real implementations have not been included here. It has also to be considered that the focus of the article is the routing protocol and not the SAS. Lastly, the performance parameters of the simulated network are unknown. As a consequence, it is not possible to evaluate, nor mitigate, eventual offsets with systems that would be physically deployed. Because of all the above-mentioned issues, the presented values should be validated by a real-world implementation before considering them reliable.

7.4. MyShake. MyShake [36] is another example of front-detection SAS based on the MEMS accelerometers embedded in common smartphones. The study identifies earthquakes that happen within a 10 km radius and that have a magnitude that is at least 5, as those that can be detected by the average smartphone. When the MyShake application is installed and active, and ad hoc algorithm continuously monitors the accelerometer and communicates shaking data to a central server if a certain condition is triggered. The central server uses a detection algorithm to confirm that an earthquake is underway and, if the output is positive, it calculates location and magnitude and issues an alarm. The paper [36] does not outline the details of the triggering algorithm on the smartphones, but it appears that the system performs an on-phone detection that is subsequently validated by a central server. The MyShake proof-of-concept has been validated by simulations, resulting in a combined delay of 5 seconds

after the origin time. Unfortunately, from the paper it is not possible to identify the delay components, although the overall figure is in line with similar systems [15, 16]. In Figure 7, data of 4 and 6 seconds for MyShake have been added for graphic purposes only.

7.5. Hualien. Hualien is a highly seismic area in Taiwan. Following a strong 6.8 M_L earthquake in 1986, the Central Weather Bureau of the country developed and tested several EEWSs [37], in a time span of 25 years. With reference to Figure 7, we present here the main outcomes of the various SASs presented in [14].

7.5.1. Hualien V0. The first deployed system, not depicted in Figure 7, was a front-detection prototype based on 10 non-MEMS force-balance accelerometer stations that continuously transmitted data to a central mainframe. During the 2-years test period, the system could provide quake warnings in approximately 10 seconds or more. As the system was also designed to provide earthquake localization and to determine quake magnitudes, its scarce reliability in these respects was enough to shut down the system.

7.5.2. Hualien V1. The subsequent iteration of the front-detection system was based on similar principles but included 110 non-MEMS force-balance accelerometer stations organized in virtual subnetworks. This is the system still currently operated by the Taiwan Central Weather Bureau. Every 10 seconds of recordings is processed by each virtual subnetwork in order to determine if an earthquake is occurring and, in the case, its magnitude and hypocenter. V1 is significantly more accurate and precise in earthquake rate detection, magnitude estimation, and hypocenter determination, at the expense of a higher average warning time. This is on average around 19*s*, although large variations exist on the reported earthquakes, ranging from 13 to 27 seconds. Times like these may be useful for locations 70 km or more away from the epicenter. While the estimation of the hypocenter can be computed in less than 10s after the arrival of the P-waves, magnitude estimation requires higher times, as it needs data from the S-waves [38]. Tests have been also made by incorporating the signals from distant sensor stations; however, it has been found that the addition in warning time is significant, for an almost negligible improvement in hypocenter and magnitude estimation [39].

7.5.3. Hualien V2. To provide warnings to locations within 70 km from the epicenter of an earthquake, an on-site prototype based on a seismographic network has been developed. Seismic signals are continuously transmitted to a central station with IP-based networks. The system tries to identify the peak magnitude of the initial P-wave displacement instead of the magnitude of the earthquake, in order to shorten processing times. Total warning time is not clear. It is claimed that this on-site alert system may offer warnings in less than 10*s*; among 54 earthquakes detected by the system, however, on average almost 19*s* have been needed to issue a warning.

7.5.4. Hualien V3. To further reduce the warning times without densifying the network with a lot of high-cost non-MEMS force-balance accelerometers, a hybrid system named *Palert*, based on MEMS accelerometers, was developed by a consortium of industry and academia. *Palert* is both an on-site and a front-detection system. Sensor devices employ a local full-fledged algorithm to detect P-waves and calculate their peak magnitudes. Sensors also send, each second, all the acceleration signals to a central server. If a P-wave is detected on-site, sensor devices start an alert with a warning sound. If the central server recognizes that a certain number of *Palert* stations are triggered, it considers it an earthquake and it starts to compute hypocenter, magnitude, and issues an alarm. For the front-detection slice, a mean value of 15.5*s*, compared to ~19*s*, and a median of approximately 14*s*, instead of ~19*s*, indicate the V3 ability to issue warnings faster, at the expense of a slightly higher uncertainty regarding magnitude and hypocenter estimation. It was concluded that *Palert* is able to function as an EEWS, for regions located at 60 km or more with respect to the quake epicenter. Regarding the on-site subsystem, it is shown how the system can issue local warnings much faster. Apart from a value of 0.01*s*, that is probably an error, the majority of stations issued warning times between 1*s* and 7*s*.

7.6. Earthcloud. On July 1, 2018, a low-power earthquake of magnitude 3.6 on the Richter scale struck the region at 07:32:16 UTC (09:32:16 local time), with a hypocenter 14 km deep and epicenter 59.11 km (36.73 mi) and 59.69 km (37.09 mi) from the first two devices and the third one, respectively, in a straight line. The farther distance where the earthquake was perceived, instead, has been reported to be approximately 58 km in a straight line from the epicenter. The Earthcloud system determined that an earthquake was underway in times between 0.7*s* and 0.8*s*. Each Earthcloud sensor, instead, detected a probable earthquake and communicated a warning in times ranging from 0.1 to 0.2 seconds. All these delays do not include magnitude estimation that is not currently performed by the prototype. The variation in time between sensor devices is due to the different network paths to the cloud system. The processing time is conservatively assumed to be ~50*ms* for each device (each device has the same processing power, and it is configured exactly the same way as the others). Transmissions between each of the two colocated devices and the AWS servers have a RTT of 106*ms*, therefore a one-way delay of 53*ms*. The same figures for the third device are instead 343*ms* and 172*ms*, respectively. As the region is the same and the number of devices is low, we considered a positive alarm only the case in which all 3 devices issued a warning in the same timeframe. The two colocated devices needed 0.275*s* to be aware that all 3 sensor nodes issued a warning, while the third device needed 0.393*s* instead. Kinesis knew of the three warnings in 172*ms*, and we assume that it may need up to 600*ms* to issue an alarm to farther sensor devices. This is confidently a worst-case scenario, as even the delays of geosynchronous satellite communications are usually lower

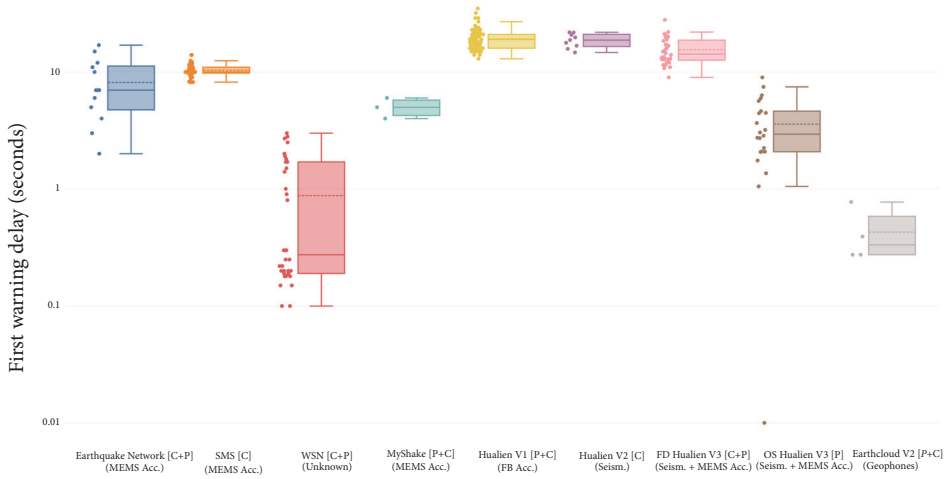


Figure 8: Comparison of first warning delays among different SASs, logarithmic scale.

[21]. Figure 8 depicts the same values of Figure 7 but on a logarithmic scale, and it is shown here to clearly compare the results of Earthcloud V2 with those of the other fastest systems.

Data harvested by Earthcloud on July 1, 2018, has been batch processed and cross-referenced with data of the same day from the Italian National Institute of Geophysics and Volcanology (INGV), which in turn sources data from the Italian National Seismic Network and other local, regional and national networks belonging to other national or international institutions. From cloud data gathered through Athena, it was confirmed that the two colocated devices detected the arrival of P-waves on all three directional axes, with peak values of $|0.5|\ mV$, at 07:32:31 UTC (09:32:31 local time) of the device clocks. The third device did not produce useful data, as it did not manage to differentiate earthquake waves from background noise. Its warning, however, contributed to the system earthquake determination. Several reasons may account for this noise in data; most likely, either there were some other vibration sources, the device was misconfigured, its assembly on the wall poorly performed, or it had some component (e.g., the analog-to-digital converter) failure. Considering a straight line distance of 59.110 km at earth level and a hypocenter with a 14.000 km depth, an approximation of the actual source point of the earthquake can be calculated as the hypotenuse of the right triangle formed through the previous figures, that is equal to 60.745 km. Counting a time difference of 15.000 s and assuming that device clocks were synchronized with INGV time, the resulting P-wave average velocity between the source and the detection point would be equal to 4.050 km/h. This figure is slightly higher than the lower end of the often-reported speed interval for P-waves of $4-8\ km/h$. Actually, as reported in [40], P-waves can travel from anywhere between 300 and 6500 m/s, depending on the terrain composition. Considering the presence of significant marshy deposits in the region (Modena, in the past, was a very swampy terrain), which can hamper wave propagation, it is plausible to expect wave propagation velocity attenuations when comparing it to average values. Furthermore, no other relevant data could be extracted from Athena in a slightly larger timeframe. Therefore, the detection was considered positive.

At 07:32:38 UTC (09:32:38 local time), the *secunda unda* was identified. Earthcloud detected S-waves until 07:32:40 UTC (09:32:40 local time), through data with peak values of $|1.125|\ mV$ from the colocated devices. Detection started at second 38 and ended at second 40, resulting in a traveling time of 22.000 s and 24.000 s, respectively, that give average velocities of 2.761 km/h and 2.531 km/h, each, calculated following the same logic used for P-waves. The mean value between the two results in 2.646 km/h. The latter is 65.333 % of the P-wave average velocity, an amount that is consistent to the typical approximate figure of 60 % that relate S-wave speed with the preceding P-wave speed, when considering the same direction and the same traversed materials. As a consequence, the detection was deemed to be positive.

7.6.1. Challenges. The number of sensor devices deployed in the Earthcloud network is still too low to definitely draw conclusions about the performance of the system. The same applies to the number of earthquakes experienced by the prototype. In particular, crucial information to extract is the relationship between the timing of definitive alerts and the accuracy of earthquake detection.

The time needed to perform the very basic processing functions in sensor devices is fixed. Being very scalable, a very similar property can be expected from the cloud. However, processing times might grow if algorithms for magnitude estimation are employed.

Due to the presence of Content Delivery Networks, middle boxes, etc., in the paths between sensor devices and cloud infrastructure, communication delays will most probably be defined by a certain range. In this regard, the most aleatory factor is instead the connection between sensor devices and the Internet. As described above, we experienced RTTs ranging from 106 to 343 milliseconds. Both include cloud processing time related to the routing of packets.

Being the infrastructure and the entry point the same, the 237*ms* difference is with certainty attributable to the network path between devices and AWS servers. The third sensor installed, in fact, instead of being directly connected to a university-grade network switch, is connected to a regular home router. As all sensors have been up to now connected to a router/switch through Ethernet cables, higher delays are expected if connecting devices through WiFi.

8. Conclusions

This article presented the second iteration of Earthcloud, a SAS designed to be low-cost, low-power, and cloud-based. Earthcloud processing and communication delays have been analyzed and compared with those of several other systems. Although more data is needed to draw solid conclusions, the Earthcloud prototype provided encouraging results. Soon after the system went operative, a small yet noticeable earthquake hit with a 14 km-deep hypocenter and an epicenter located approximately 60 km in a straight line from where sensor devices were deployed. Earthcloud emitted a number of warnings, validated through the collected data that indicate a significant correlation between the elaborated material and the seismic data published by the relevant national authority. All warnings issued by the prototype have been signaled in less than 1 second. This delay is in stark contrast with traditional SASs, as well as with later low-cost alternatives based on MEMS accelerometers and local processing; in both cases, these systems usually give alerts in approximately 5 seconds when on-site, or in 10 to 20 seconds when front-detection. Part of the reason is that they usually cover a very specific area while employing some form of magnitude and epicenter location estimation. While the latter would be, in perspective, given also by the Earthcloud system as it is, the former would need the employment of additional processing algorithms that would presumably increase the prototype warning times.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to express their gratitude to both the Department of Engineering Enzo Ferrari and the Ovestlab Center (http://ovestlab.it) together with its head, Silvia Tagliazucchi, for allowing us to use their respective premises and resources for our devices without charge.

References

[1] R. R. Leach Jr. and F. U. Dowla, "Earthquake early warning system using real-time signal processing," in *Proceedings of the IEEE Signal Processing Society Workshop*, pp. 463–472, September 1996.

[2] R. Hoque, S. Hassan, M. A. Sadaf, A. Galib, and T. F. Karim, "Earthquake monitoring and warning system," in *Proceedings of the 3rd International Conference on Advances in Electrical Engineering, ICAEE 2015*, pp. 109–112, Bangladesh, December 2015.

[3] Y. Sherki, N. Gaikwad, J. Chandle, and A. Kulkarni, "Design of real time sensor system for detection and processing of seismic waves for earthquake early warning system," in *Proceedings of the IEEE International Conference on Power and Advanced Control Engineering, ICPACE 2015*, pp. 285–289, India, August 2015.

[4] T. Uga, T. Nagaosa, and D. Kawashima, "An emergency earthquake warning system using mobile terminals with a built-in accelerometer," in *Proceedings of the 2012 12th International Conference on ITS Telecommunications, ITST 2012*, pp. 837–842, Taiwan, November 2012.

[5] J. Nachtigall, A. Zubow, R. Sombrutzki, and M. Picozzi, "The challenges of using wireless mesh networks for earthquake early warning systems," in *Proceedings of the 2009 2nd International Conference on Advances in Mesh Networks, MESH 2009*, pp. 155–162, Greece, June 2009.

[6] J. Nachtigall and J.-P. Redlich, "Wireless alarming and routing protocol for Earthquake Early Warning Systems," in *Proceedings of the 4th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2011*, pp. 1–6, France, February 2011.

[7] A. Heryana, E. Nugraheni, B. Kusumo, A. F. Rojie, and B. Setiadi, "Applying agile methods in designing an earthquake and landslide early warning system application for Android," in *Proceedings of the 5th International Conference on Computer, Control, Informatics and its Applications, IC3INA 2017*, pp. 80–84, Indonesia, October 2017.

[8] E. Shayo, A. Mwambela, and S. Kawambwa, "Evaluating quality of service for short message service as a platform to transmit seismic data in earthquake early warning system," in *Proceedings of the 2018 IST-Africa Week Conference, IST-Africa 2018*, Botswana, May 2018.

[9] E. Heindl, "Peer-to-peer (P2P) earthquake warning system based on collaborative sensing," in *Proceedings of the 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST '09*, pp. 174–176, Turkey, June 2009.

[10] F. Finazzi, "The earthquake network project: toward a crowdsourced smartphone-based earthquake early warning systemthe earthquake network project: toward a crowdsourced smartphone-based earthquake early warning system," *Bulletin of the Seismological Society of America*, vol. 106, no. 3, p. 1088, 2016.

[11] F. Finazzi and A. Fassò, "A statistical approach to crowdsourced smartphone-based earthquake early warning systems," *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 7, pp. 1649–1658, 2017.

[12] A. Zaicenco, S. Huffman, and I. Weir-Jones, "Seismic P-wave polarization in the context of on-site early warning system," in *Proceedings of International Conferences on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics*, May 2010.

[13] "Earthquake and Tsunamis - disaster prevention and mitigation efforts," Japan Metereological Agency, March 2018, http://www.jma.go.jp/jma/kishou/books/jishintsunami/en/jishintsunami_en.pdf.

[14] Y.-M. Wu, N.-C. Hsiao, T.-L. Chin, D.-Y. Chen, Y.-T. Chan, and K.-S. Wang, "Earthquake early warning system in taiwan," in *Encyclopedia of Earthquake Engineering*, pp. 1–12, Springer, Berlin, Germany, 2013.

[15] E. R. Burkett, D. D. Given, and L. M. Jones, "ShakeAlert - an earthquake early warning system for the United States west coast," U.S. Geological Survey Open-File Report 2014–1097, USGS Earthquake Science Center, February 2017.

[16] D. Given, E. Cochran, T. Heaton et al., "Technical implementation plan for the ShakeAlert production system - an earthquake early warning system for the west coast of the united states," U.S. Geological Survey Fact Sheet 2014–3083, USGS Earthquake Science Center, May 2014.

[17] M. Klapez, C. A. Grazia, S. Zennaro, M. Cozzani, and M. Casoni, "First experiences with earthcloud, a low-cost, cloud-based iot seismic alert system," in *Proceedings of the 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 262–269, Limassol, Cyprus, October 2018.

[18] "Power: what are the power requirements?" Web page, Raspberry PI Foundation, 2019, https://www.raspberrypi.org/documentation/faqs/#pi-power.

[19] K. Faber and P. W. Maxwell, "Geophone spurious frequency: what is it and how does it affect seismic data?" in *Proceedings of the SEG Technical Program Expanded Abstracts 1996*, pp. 79-80, 2005.

[20] M. Peter and M. Lansley, "What receivers will we use for low frequencies?" in *Proceedings of the SEG San Antonio Annual Meeting*, pp. 72–76, 2011.

[21] M. Casoni, C. A. Grazia, M. Klapez, N. Patriciello, A. Amditis, and E. Sdongos, "Integration of satellite and LTE for disaster recovery," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 47–53, 2015.

[22] C. A. Grazia, M. Klapez, N. Patriciello et al., "Integration between terrestrial and satellite networks: The PPDR-TC vision," in *Proceedings of the 2014 10th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2014*, pp. 77–84, Cyprus, October 2014.

[23] M. Klapez, C. A. Grazia, and M. Casoni, "Towards massively multipath transmissions for public safety communications," in *Proceedings of the 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–7, New York, NY, USA, October 2016.

[24] M. Casoni, C. A. Grazia, and M. Klapez, "A software-defined 5G cellular network with links virtually pooled for public safety operators," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 3, p. e3092, 2017.

[25] M. Casoni, C. A. Grazia, and M. Klapez, "SDN-based resource pooling to provide transparent multi-path communications," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 172–178, 2017.

[26] M. Klapez, C. A. Grazia, and M. Casoni, "A hybrid algorithm to combine redundancy and concurrency in virtual network resource pooling," *Internet Technology Letters*, vol. 1, no. 2, p. e18, 2018.

[27] C. A. Grazia, M. Klapez, N. Patriciello, and M. Casoni, "Pink: Proactive injection into ack, a queue manager to impose fair resource allocation among tcp flows," in *Proceedings of the 11th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2015*, pp. 132–137, UAE, October 2015.

[28] M. Casoni, C. A. Grazia, M. Klapez, and N. Patriciello, "How to avoid TCP congestion without dropping packets: an effective AQM called PINK," *Computer Communications*, vol. 103, supplement C, pp. 49–60, 2017.

[29] A. A. Gebremariam, M. Usman, R. Bassoli, and F. Granelli, "SoftPSN: software-defined resource slicing for low-latency reliable public safety networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7253283, 7 pages, 2018.

[30] M. Casoni, C. Augusto Grazia, and P. Valente, "Achieving a high throughput and a low latency through a modular packet scheduler," *International Journal of Communication Networks and Distributed Systems*, vol. 20, no. 1, pp. 82–109, 2018.

[31] M. Casoni, C. A. Grazia, M. Klapez, and N. Patriciello, "A congestion control middleware layer with dynamic bandwidth management for satellite communications," *International Journal of Satellite Communications and Networking*, vol. 34, no. 6, pp. 739–758, 2016.

[32] M. Casoni, C. A. Grazia, M. Klapez, and N. Patriciello, "QRM: a queue rate management for fairness and TCP flooding protection in mission-critical networks," *Computer Networks*, vol. 93, pp. 54–65, 2015.

[33] N. A. Surobhi and A. Jamalipour, "A MANET-based semantic traffic management framework for ubiquitous public safety networks," *Wireless Communications and Mobile Computing*, vol. 14, no. 12, pp. 1127–1142, 2014.

[34] D. Giardini, G. Grünthal, K. M. Shedlock, and P. Zhang, "The gshap global seismic hazard map," *Annals of Geophysics*, vol. 42, no. 6, 1999.

[35] L. Arcoraci, M. Berardi, F. Bernardini et al., "Rapporto macrosismico sui terremoti del 20 (ML 5.9) e del 29 MAGGIO 2012 (ML 5.8 E 5.3) nella pianura padano-emiliana," Official Report, Istituto Nazionale di Geofisica e Vulcanologia, July 2012, http://terremoti.ingv.it/images/pdf/QUEST_Emilia2012_RapportoFinale.pdf.

[36] Q. Kong, R. M. Allen, L. Schreier, and Y. Kwon, "MyShake: a smartphone seismic network for earthquake early warning and beyond," *Science Advances*, vol. 2, no. 2, Article ID e1501055, 2016.

[37] Y.-M. Wu, N.-C. Hsiao, T.-L. Chin, D.-Y. Chen, Y.-T. Chan, and K.-S. Wang, "Earthquake early warning system in Taiwan," in *Encyclopedia of Earthquake Engineering*, pp. 1–12, Springer, Berlin, Germany, 2014.

[38] Y.-M. Wu and H. Kanamori, "Experiment on an onsite early warning method for the Taiwan early warning system," *Bulletin of the Seismological Society of America*, vol. 95, no. 1, pp. 347–353, 2005.

[39] Y.-M. Wu and T.-L. Teng, "A virtual subnetwork approach to earthquake early warning," *Bulletin of the Seismological Society of America*, vol. 92, no. 5, pp. 2008–2018, 2002.

[40] O. Coussy and B. Zinszner, *Acoustics of Porous Media*, Editions Technip, 1987.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 1408784, 10 pages
https://doi.org/10.1155/2019/1408784

Research Article

Next Generation Emergency Services Based on the Pan-European Mobile Emergency Application (PEMEA) Protocol: Leveraging Mobile Positioning and Context Information

Urban Sedlar,[1] **James Winterbottom**,[2] **Bostjan Tavcar**,[3] **Janez Sterle**,[4] **Jaka Cijan**,[1] **and Mojca Volk**[1]

[1] *University of Ljubljana, Slovenia*
[2] *Deveryware S.A., France*
[3] *Administration of the Republic of Slovenia for Civil Protection and Disaster Relief, Slovenia*
[4] *Internet Institute, Ltd., Slovenia*

Correspondence should be addressed to Urban Sedlar; urban.sedlar@fe.uni-lj.si

Received 30 October 2018; Revised 4 February 2019; Accepted 5 March 2019; Published 24 March 2019

Guest Editor: Maurizio Casoni

In this paper, we analyze requirements of next generation 112 emergency services in the era of ubiquitous mobile devices and sensors and present the design, implementation, and piloting results of our testbed, which was developed within the H2020 project NEXES. The system leverages a multihop location-aware PEMEA routing network that finds the geographically closest responsible public service answering point (PSAP) and supports cross-border application roaming. Our reference mobile implementation utilizes multiple device and network-based positioning technologies, which, combined, both outperform traditional cell-tower based positioning and provide a means for detecting fraudulent calls. The system is extensible and can establish a variety of communication channels after the initial emergency session is set up; we demonstrate this with an interoperable WebRTC-based video call. The obtained results demonstrate the viability and flexibility of PEMEA-based over-the-top emergency services, show high user acceptance when comparing them with existing solutions, and thus pave the road for further rollout of such systems.

1. Introduction

In recent years, ubiquitous sensing and positioning have enhanced several applications for general population. For example, web search engines commonly tailor results to user's location; nearby events and stores are promoted on social networks and through advertising platforms. Commercial applications even extend to indoor positioning based on advanced Wi-Fi network models and Bluetooth beacons [1].

However, emergency services throughout the world are still largely relying on network-based mobile phone tracking, which either utilizes cell identification, cell coverage, triangulation, or trilateration. Such approaches yield accuracy in the range of a couple of hundred meters at best, which is much worse than the previous landline number-to-location mapping. For improved emergency services a better location is needed, which has to be provided either by a denser deployment of base stations [2] or by other sensors or mechanisms, implemented in the handset itself [3–5].

The latter option is already viable today; off-the-shelf mobile handsets commonly use a Global Navigation Satellite System (GNSS) receiver [6] and can look up the location of nearby Wi-Fi access points [7] and cell IDs in large crowdsourced databases provided by companies such as Apple, Google, or Combain. General-purpose emergency call solutions that leverage such handset-based positioning today are either SMS- or HTTP-based or encompass proprietary

applications that are tied to a specific PSAP (e.g., a national mobile application tied to a national PSAP).

In this paper, we describe how we deployed and tested an end-to-end system for data-based emergency calls based on a new roaming-capable standard for Pan-European Mobile Emergency Application (PEMEA) [8]. The described testbed comprises a mobile application with a backend server, PEMEA network nodes, a PSAP system, and multiple over-the-top and legacy communication services, ranging from a PSTN and SIP calls to a WebRTC video call.

In Section 2 we examine the related work in this area. Section 3 outlines the requirements that we elicited from various system stakeholders. Section 4 describes system design and Section 5 the implementation details. In Section 6, we evaluate the system from technical and user perspectives. In Section 7, we provide concluding remarks.

2. Related Work

Multiple initiatives and projects have been trying to address the problem of contacting emergency services and providing better information faster [9]. The NG112 Long Term Definition [10] document by the European Emergency Number Association (EENA) focuses primarily on bringing together today's heterogeneous telecommunications systems, such as legacy telephony networks, telco-managed IP-based voice services (such as Voice over LTE), and a dedicated Emergency Services Internet Protocol Network (ESInet). Most of these systems are telco voice centric and will be interconnected using gateways and a number of specialized functional elements for call routing, media transcoding, location information exchange, and similar. As a vision for the future, integration with over-the-top (OTT) systems is also foreseen, such as web-based solutions, but these are currently nonexistent in practice and only recently have standards begun to emerge. Various research projects are extending this space in the directions of accessibility for the elderly and people with disabilities, social media integration, haptics, automated calling, and similar [11–14].

Rudimentary data exchange, which is possible in such systems, already provides a range of possible benefits over voice calls, starting with more efficient and less error-prone data collection when every second counts [3, 15, 16], automated triggering as it happens in the vehicular domain with eCall [17–19], to the possibility of automated triage on the PSAP side in case of large-scale events [20].

The actual mechanics of data delivery are usually solved on a national level, with integrations between operators and PSAPs, either by event-based message delivery or by the use of location and information platforms with a central database (e.g., the Polish PLI CBD).

A notable and more recent standard in the mobile domain is the Advanced Mobile Location (AML) (ETSI Technical Report (TR) EMTEL-00035), which is considered as an enhancement of legacy emergency systems. Today, AML is supported in majority of mobile devices on the market (Android and iOS based), but only a handful of countries currently support it. AML typically uses an invisible-to-the-user text or binary SMS [21] to send the emergency-related data when an emergency number is dialed. It can be configured to post the data to a REST API as well.

Other innovative solutions in this space include using text-to-speech to relay preset information and location information to the dispatcher automatically using voice, as well as using proprietary information repositories to store such information. Another possibility is the E.164 Number to URI Mapping (ENUM) standard that maps phone number identities to records in the Domain Name System (DNS); this is already implemented in some countries and could be used for the discovery of a Location Information Server (LIS), storing the user's location. In our opinion, these solutions are a poor fit in terms of both required flexibility and regulations.

In addition, none of these solutions has adequately addressed the aspects of international roaming. This is relevant not only due to the increasing number of travelers, but also because it ultimately enables mobile application interchangeability. A small PSAP can thus leverage a higher-quality app developed by another region or country. A proposal addressing this problem is the recently published PEMEA standard [8, 22, 23], which is also an important first step towards the interconnection of the telco emergency infrastructure systems with the OTT world. In this paper, we try to demonstrate the viability of PEMEA in practical scenarios. Further, we aim to demonstrate that it can be extended with support for various communication technologies; we focus specifically on web-based OTT technologies that can provide an emergency services ecosystem that is secure and reliable and at the same time sufficiently open for new developers and services.

3. Requirements Engineering

We defined our design guidelines and constraints through the process of requirements engineering, with the help of a number of real world emergency systems stakeholders, including 112 PSAP operators, first responders, third-party solution developers, and equipment vendors. During the process, we have organized multiple workshops and tabletop exercises for collection of requirements, as well as feedback, both before and during the system implementation. Besides the usual requirements of reliable operation, user-friendliness, providing adequate feedback to the caller, support for GNSS positioning, and support for video communication to cater especially to the deaf and hard of hearing people, we also identified multiple less obvious requirements and features:

(i) Positioning using multiple technologies (both device based: GNSS, Wi-Fi, and mobile network based)

(ii) Medical data transfer with adequate data privacy: on-demand transfer of data only to the ambulance service dispatchers, bypassing intermediaries

(iii) Ability to capture as much contextual information as possible (battery level, network type, data downlink, and uplink speeds) to guide further communication decisions and constraints

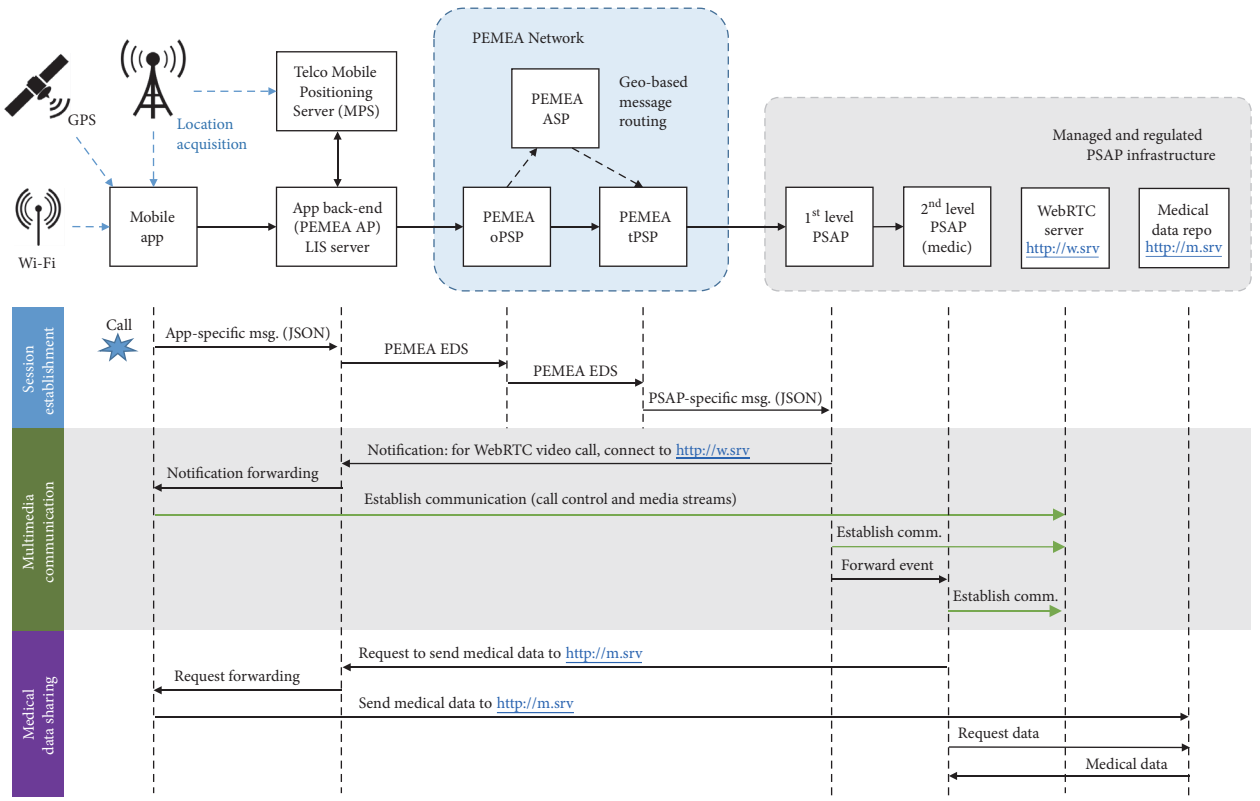


Figure 1: Pilot system architecture and main data flows. Green arrows denote media (audio/video) streams.

(iv) International roaming, to enable an app connected to its own backend to be used abroad

(v) Effortless extensibility with PSAP's own content and web-based services, such as web-based surveys

(vi) Chats, media (images and uploaded videos) should be stored on the premises of the destination PSAP

(vii) All audio and video communication should be recorded and stored on the premises of the destination PSAP.

Based on these requirements, the system design was iteratively refined until the requirements were met and most of the existing and desired procedures could be executed from beginning to end.

4. System Design

The final end-to-end architecture of the system is presented in Figure 1, together with three main groups of message flows. In the following sections, we describe each subset of the architecture.

4.1. PEMEA Network. The core of the architecture is shown in a blue block (Figure 1). PEMEA architecture consists of basic routing elements, PEMEA Service Provider (PSP) components. Each PSP can be either an entry node for emergency messages (termed originating PSP or oPSP), or an exit node that talks to a PSAP (in this case it is called a terminating PSP or tPSP). An identical node can operate at a higher level and forward messages between PSPs; then it becomes an Aggregating Service Provider (ASP). A message with a predefined format is passed between the PSPs and ASPs, called the *EmergencyDataSend* (EDS) message. EDS is defined with an XML schema in the PEMEA specification [22] and includes geographic coordinates of the sender, as well as some routing-related information (time to live, list of hops, etc.). All information passed between PEMEA nodes is communicated through SSL REST APIs that are mutually authenticated using X.509 certificates.

To secure the system, PEMEA expects all entry and exit nodes to trust their sources and destinations, respectively, by using a Fully Qualified Doman Name (FQDN) whitelist.

4.2. Data Source and Destination. The sources of emergency messages in the EDS format are called Application Providers (AP). In most cases, a source is expected to be a mobile app backend (relaying data of all of its users), rather than an individual mobile app. Mutual authentication and whitelisting mechanisms make it easy to block a rogue backend that is sending out a large amount of fraudulent calls. To store the locations, a Location Information Server (LIS) can be used, which was in our case collocated with the AP.

On the other side of the network, an exit node is expected to be a PSAP. There are no data format requirements here, and integration is expected to be done on a case-by-case basis with existing proprietary PSAP software. Additionally,

PSAPs typically operate in different hierarchies, according to national specifics. The received data can be forwarded between PSAPs, making it possible to share the same set of data and communication channels.

4.3. Multimedia Communication. The PEMEA system in itself serves only to establish a basic session. No multimedia or additional communication is sent through the PEMEA nodes. Rather, the source AP provides a set of callback Uniform Resource Identifiers (URIs) that can be used to establish multimedia streams, trigger calls, and exchange ancillary data. Figure 1 shows how a media session establishment by means of a callback URI takes place; the PSAP notifies the AP directly that a Web Real-Time Communication (WebRTC) session is taking place in a WebRTC web application on a provided Uniform Resource Locator (URL) address. Once both the mobile application and the PSAP console open the same web view, a communication session is established using web-based technologies. This makes it easy to deploy not only real-time multimedia communication, but also text chat, or any kind of web-based collaboration. Other actors (such as 2nd level PSAPs, medics, and firemen) can also participate in the discussion by opening a unique URL of the chat room with a temporary session token. Since any actual communication service is hosted on a managed and regulated infrastructure of the PSAP, it also allows session recording and largely solves the problems of data management, data retention, and auditing.

4.4. Sensitive Data Exchange. Exchange of any kind of data after the delivery of the initial PEMEA message happens directly between the PSAP and the AP. However, to prevent even the AP from having access to the sensitive user-related data (photos, media, medical information, and similar), a more secure mechanism was implemented, as evident from Figure 1, as follows: a request for sensitive data is sent to the AP and then relayed to the mobile app. This request includes an endpoint under the control of the PSAP, where the data can be safely deposited. In the next step, the app asks the user for permission to deposit private data, and, if permission is given, the data is sent directly from the mobile app to the PSAP-controlled endpoint.

5. Pilot Implementation

The following sections present details about our pilot implementation that was undertaken with the end goal of being able to evaluate the system in a large demonstration event with real-world stakeholders.

5.1. Mobile Application. The Android mobile application features a large SOS/112 (112 is the standard European emergency number, equivalent to 911 in the US. More information at http://www.eena.org/) button to start a call. In addition, the app includes a detailed profile form where the citizen can beforehand provide all relevant contact or medical information (see Figure 3). The data model for this was inspired by Apple's Medical ID on iOS, a feature that also hints how the future emergency services apps could look when deeply integrated into the mobile OS.

5.2. Mobile App Backend. The app backend (AP) serves as a link to the PEMEA network and as a relay to reach the mobile app from the outside world when it is behind a NAT/Firewall. In addition, we integrated it with the telco infrastructure to obtain another location data point from the mobile network itself. This data can either validate or invalidate the location reported by the phone. For each emergency session, the backend server generates multiple tokens and sets up multiple proxy URLs, such that all communication sent to the proxy URLs is relayed back to the mobile app. Finally, the backend assembles the PEMEA *EmergencyDataSend* message, which it then forwards to the *originating* PEMEA Service Provider node (oPSP).

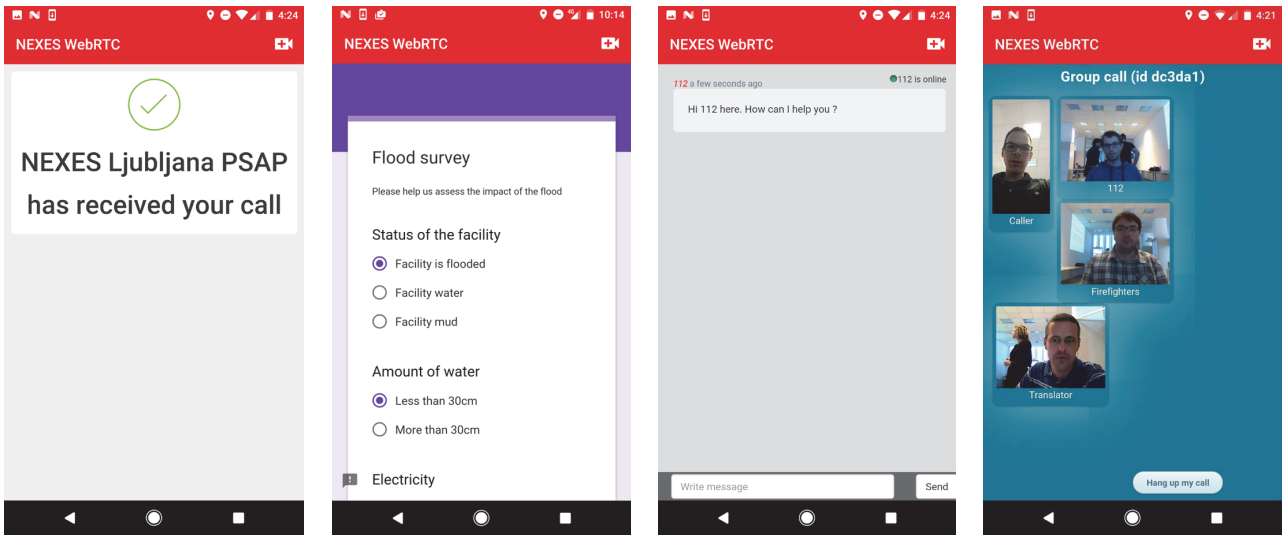
5.3. PEMEA PSP with GIS-Based Message Routing. PEMEA PSP element was developed from scratch according to the specification [22] and tested during interoperability tests with the NEXES [14] project partners. The PSP element receives the incoming EDS message, parses out the location, and forwards the message, to either a known endpoint, another PSP, or an ASP element (default gateway) when no better route is found. The PSP was set up to serve 13 regions in Slovenia (Figure 4) based on their *shapefiles* and a PostGIS query was used to determine if a location falls within any of the regions.

5.4. PSAP. To receive the calls and initiate further communication, we developed a simplified PSAP console with a Hyper-Text Markup language (HTML) based frontend. The use of web technologies made it easier to upgrade the established emergency sessions to other web-based channels, such as HTML information display, WebSockets-based chat, and WebRTC video communication.

5.5. Multimedia: WebRTC Server. Multimedia communication was implemented through WebRTC, an HTML5-based real-time communication stack that is available in most modern web browsers. We used WebRTC for its ease of creating multiparty video chatrooms and the requirement of simplicity to join such calls on the side of the first responders; in case of WebRTC, only a modern Web browser is needed.

However, WebRTC is by design peer-to-peer, which means that media cannot be easily recorded, even when required by regulations. Thus, we used a slightly modified architecture, with a server placed in lieu of a peer; the server then acts as a central point of communication, providing audio/video mixing and recording. A mobile app screen with an active video chat room based on the open-source Kurento WebRTC server is shown in Figure 2(d).

5.6. Other Web-Based Services. By using a flexible web view, many more modes of communication can be implemented using the standard popular web development stacks. A static HTML-based notification is shown in Figure 2(a); in non-critical situations, interactive apps such as web-based surveys can help with crowdsourced data collection (Figure 2(b)).



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

FIGURE 2: Full-featured Chromium browser inside the mobile application, for hosting different extensions. From left to right: (a) web-based notification; (b) a survey; (c) WebSocket-based chat application; (d) WebRTC-based multiparty video.

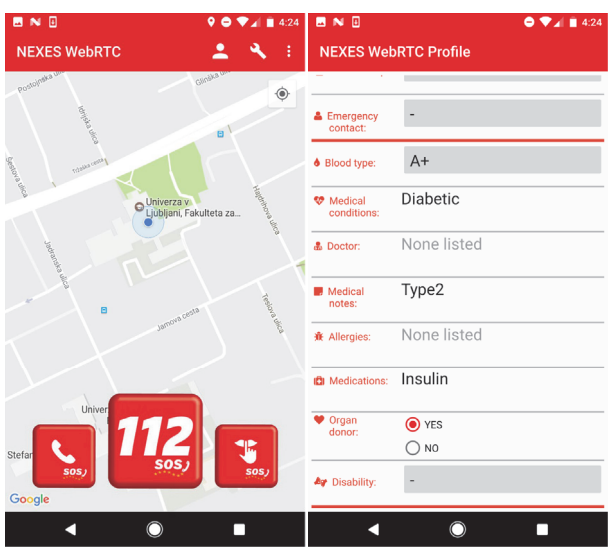


FIGURE 3: Citizen Application for Android: main screen with location display and emergency call button (left); user information screen (right).

Similarly, we developed a web-based chat, leveraging WebSockets for communication. The mobile side of the app can be seen in Figure 2(c). But the high pace of emerging web technologies already makes it possible to develop and deploy Web Graphics Library (WebGL) based visualizations and Web Virtual Reality (WebVR) or Web Augmented Reality (WebXR) scenes that could in the future help users by displaying virtual reality worlds or overlays.

6. Evaluation

Evaluation was performed at our public demonstration event with participants and observers from all major stakeholder groups (citizens, including special needs minorities, as well as national police, fire brigade, ambulance service, municipality disaster relief team, equipment vendors, and telcos).

The example depicted in Figure 5 shows a deaf person who requires a sign language interpreter for communication. Once the session is established, the 112 operator is able to see the profile of the "caller" and their disability. The operator can thus decide to establish a video call and then patches a sign language interpreter into the multiparty video call to help with communication. Once the information is received, the caller can remain online as the 2nd-level fire department PSAP joins in on the same video call. All communication is relayed through the central WebRTC server located with the PSAP, which also performs recording. Such setup allows easy participation of any required stakeholder, in most cases without the need for special applications; all that is required is a modern web browser and the knowledge of the temporary session token identifying the video chat room.

In addition, four more scenarios were successfully demonstrated and role-played:

(i) A mudslide event: uploaded pictures and video clips help emergency services identify scope of the event. The 112 PSAP and a 2nd-level fire department PSAP are involved.

(ii) A lost person event: The 112 PSAP and a 2nd-level ambulance service PSAP are involved and find the citizen through the periodic location updates.

(iii) A traffic accident of a tourist: the call is relayed through tourist's home country and the PEMEA network, arriving to the geographically closest PSAP. The 112 PSAP and a 2nd-level police PSAP are involved.

(iv) A hiking accident: a diabetic citizen cannot explain what is going on, but a preauthorized medical data

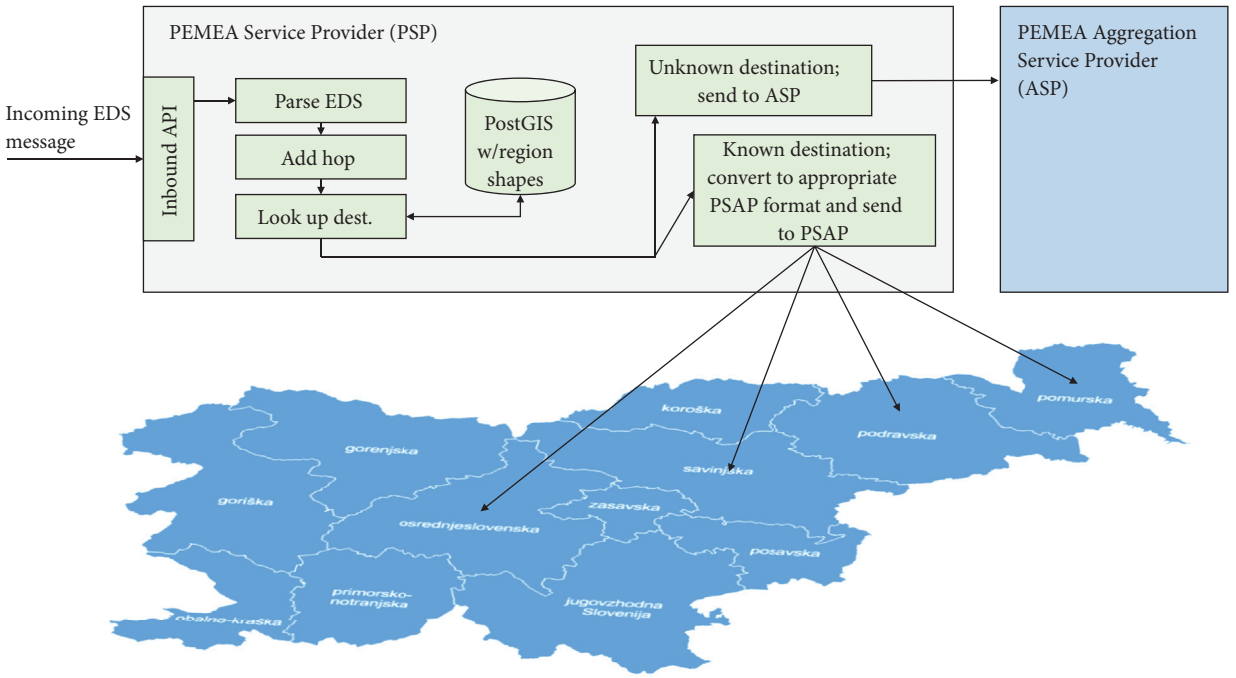


Figure 4: PSP architecture with PostGIS shapes for regions in Slovenia. Messages with served location are transformed to PSAP-specific format and sent to appropriate PSAP.

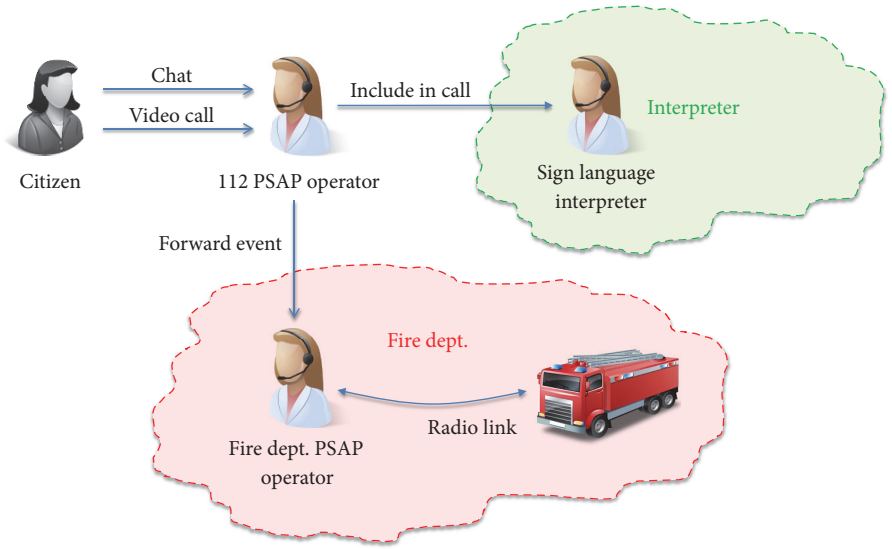


Figure 5: Example scenario of a deaf person calling 112 services; a sign language interpreter is added to the video call to help with communication.

request helps the 112 PSAP to determine this is a medical emergency.

6.1. User Acceptance and Feedback. After the demonstration event, the participants were able to experiment with the solution and try out the roles on the side of the citizen (mobile app) as well as on the side of the PSAP and emergency response organizations (ERO). Of over 60 participants, 25 were willing to provide written feedback by filling out one of two types of surveys (from the perspective of either PSAP/ERO, the citizen, or a deaf citizen), rating the improvement of the current situation on a 5-point Likert scale. All three cohorts agreed that the presented features improved upon the existing systems; a detailed breakdown of the questions and the responses is presented in Figure 7.

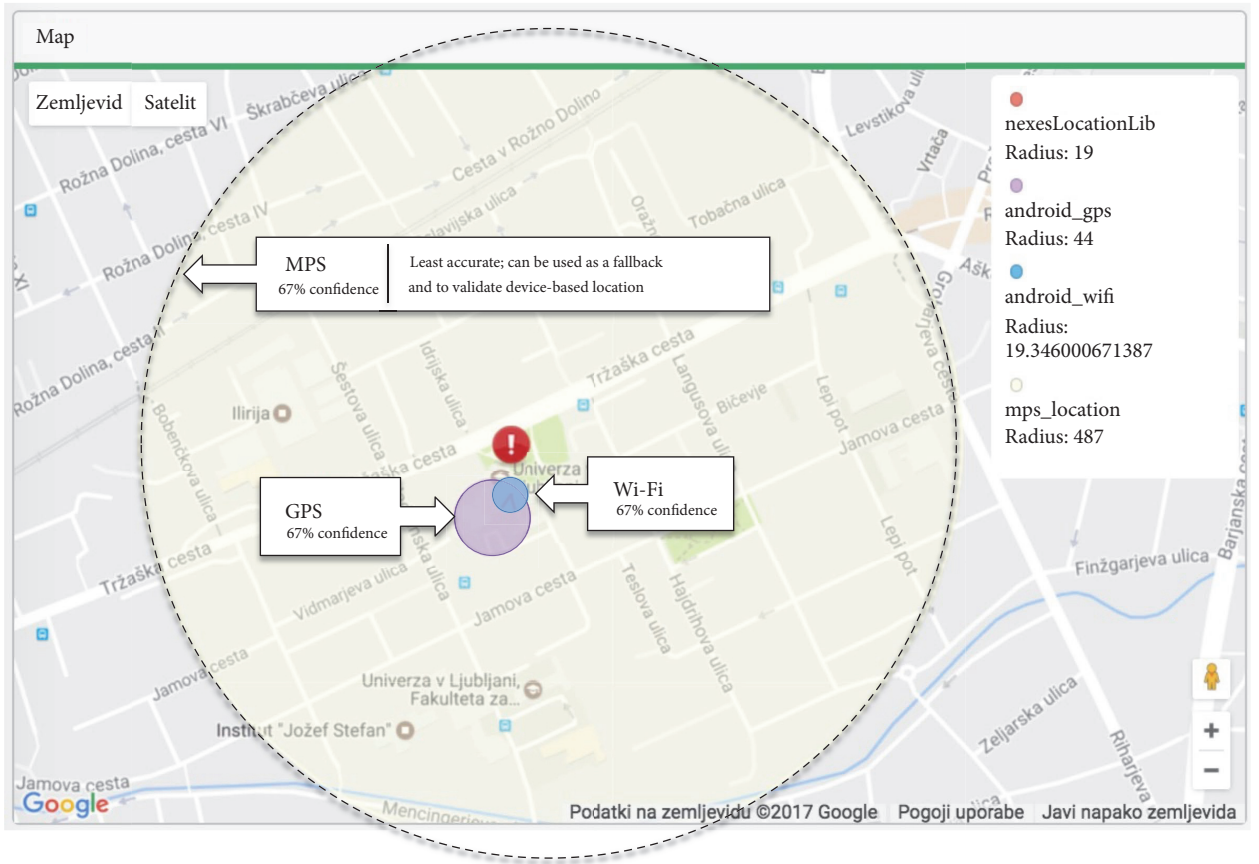


Figure 6: PSAP dashboard screenshot, demonstrating all three obtained locations: GPS, Wi-Fi, and telco network-based MPS.

In addition, during the subsequent discussions, the professional cohort expressed that the location feature was most important, followed by media sharing, and, lastly, due to a small percentage of international incidents, roaming. The citizens cohort ranked the importance from better location (most important), chat, international roaming, to video (least important). The deaf cohort ranked video and text chat as the most important features.

6.2. Positioning Accuracy. The following differences between client-based positioning technologies (GPS and Wi-Fi) and an existing telco Mobile Positioning System (MPS) were determined, based on 400 test calls in different locations around our campus (see also Figure 6) (see Table 1).

Best GPS improvement over the actual MPS result was determined as 442x smaller radius (outdoors), yielding 195000x smaller search area; worst improvement of GPS over MPS was only a 2x smaller radius (4x smaller search area). Additionally, best improvement of Wi-Fi over MPS was 148x (22000x smaller search area), but Wi-Fi also performed worse than MPS in 24% of our test cases, which proves it is not very reliable but can still provide valuable insight in majority of the cases.

6.3. International Roaming Capabilities. One of the goals of the pilot was to demonstrate international roaming of an emergency data exchange (emergency session establishment), which was achieved by leveraging the PEMEA testbed routing capabilities.

For the purpose of the pilot, every emergency session establishment was performed using the PEMEA *Emergency-DataSend* message; in case of national scenarios they were performed through only one, national, PSP element that served as both oPSP and tPSP at the same time. In these scenarios, the main task of the PSP was resolution of local PSAPs based on Slovenian region boundaries.

However, in the international roaming scenario, a foreign citizen used their own national (Spanish) 112 mobile application. This application was by design only able to talk with its own backend in Spain, which had no mapping of Slovenian regions. Spanish backend thus sent data to its only known PSP in Spain, and over a number of hops the PEMEA network elements routed the message to the appropriate PSAP in Slovenia. This is shown in Figure 8.

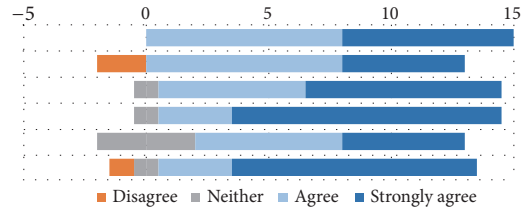
6.4. System Overhead. This section presents the application power consumption, CPU utilization, and overall system delays of the pilot solution. Three consumer mobile devices were used for analysis of the system overhead: HTC 10 running Android 6, Sony Xperia XZ running Android 8, and Google Pixel 3 running Android 9.

Table 1

	GPS reported accuracy (radius)	Wi-Fi reported accuracy (radius)	MPS reported accuracy (radius)
Min	1 m (outdoors)	3.6 m	440 m
Max	245 m (indoors)	1700 m	592 m
Mean	38.9 m (all situations)	452 m (all situations)	480 m
Median	15 m (all situations)	37.5 m (all situations)	487 m

Does the demonstrated solution improve current PSAP/Emergency response organization (ERO) capabilities?

PSAP/ERO feedback (N=15)	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Overall	0	0	0	8	7
Text chat only	0	2	0	8	5
Media transfer only	0	0	1	6	8
Video call only	0	0	1	3	11
International call routing	0	0	4	6	5
Improved location	0	1	1	3	10



Does the demonstrated solution improve current capabilities of emergency calling?

Citizen feedback (N=6)	Strongly disagree	Disagree	Neither	Agree	Strongly agree
Overall	0	0	0	2	4
Text chat only	0	0	1	3	2
Media transfer only	0	0	0	3	3
Video call only	0	0	0	3	3
International call routing	0	0	1	2	3
Improved location	0	0	1	2	3
Deaf association citizen feedback (N=4)					
Overall	0	0	0	2	2
Text chat only	0	0	1	1	2
Media transfer only	0	0	1	1	2
Video call only	0	0	1	1	2
International call routing	0	0	2	2	0
Improved location	0	0	1	2	1

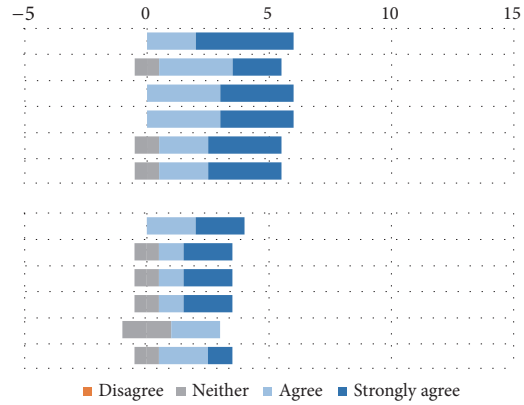


Figure 7: PSAP/emergency response organization and citizen feedback regarding overall improvements of the proposed system compared to existing solutions.

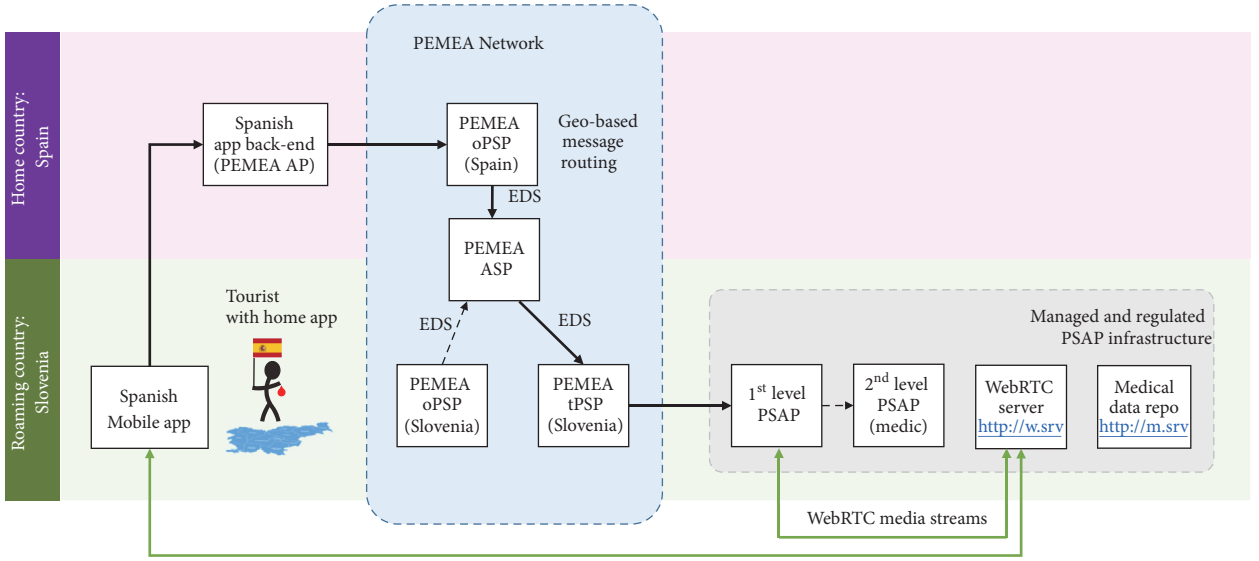


Figure 8: Roaming scenario architecture; thick black arrows indicate the path of the EDS message; green arrows indicate WebRTC media streams.

6.4.1. Mobile App CPU Usage and Power Consumption. To estimate CPU usage, Android Debug Bridge was used, with the command *adb shell dumpsys cpuinfo*. For power consumption estimation, we used Digibites AccuBattery, which logs the battery discharge rate reported by the hardware. All measurements were done with the app in foreground and screen brightness set to 80%.

The CPU usage of the citizen mobile app is low in idle state, that is, less than 1%. Similarly, the estimated power usage is negligible, even though the app activates GNSS reception immediately at start-up, to be able to acquire a satellite fix by the time user initiates a call. GNSS does contribute to the power consumption, but since that is the most important information to relay to the PSAP, it is enabled regardless of the battery level.

The majority of the battery drain comes from the CPU usage if a video call is established. CPU usage during WebRTC video call ranges from 80% on newer Pixel 3 to over 120% (more than 1 core fully utilized) on both older models. The estimated total current draw ranges between 850 mA and 1150 mA across the 3 hardware models, which is in line with the observed discharge rate of between 30% and 42% of total battery capacity per hour during a 1-hour WebRTC call.

When the user establishes an emergency session, we also profile the connectivity using Internet Control Message Protocol (ICMP) ping and perform a short downlink and uplink bandwidth test by transferring a small file. The main purpose of this test is to be able to determine if video call is a viable alternative for reaching the user later. Due to the battery drain of video calls, we omit both bandwidth profiling and video call establishment if battery level is below 40%.

6.4.2. System Delays. The delays in the system come from the following parts: (i) initial request transmission, which depends on mobile network latency and bandwidth, (ii) application server (AP) processing time (with added overhead of MPS resolution delay), (iii) PSP node processing (message parsing, validation, and routing) at each hop (multiplied by the number of hops), and, finally, (iv) the PSAP server processing time. Using only one PSP hop between the AP and the PSAP, we measured the average end-to-end message delay in urban environments (4G network) at 1.87 seconds, which includes the network profiling tests. With network profiling disabled, the average end-to-end message delay on a 4G network was 1.36 s from the mobile app to the PSAP dashboard. In contrast, with 3G network connectivity, the average end-to-end message delay was 2.2 s, and, with only 2G connectivity, the average delay increased to 4.90 s without network profiling, which was automatically disabled in both latter cases.

Another significant source of delays was connected to the WebRTC video call establishment, which was enabled on 4G networks only. This took on average 5.8s, which includes pushing the WebRTC app URL to the device, downloading a client-side WebRTC HTML5 app to the phone, assembling a Session Description Protocol (SDP) message, transferring it, and finally establishing a media stream.

7. Conclusions

In this article we presented the design of a next generation 112 emergency system based on the PEMEA protocol. The solution has been designed and implemented from scratch by the partners of the H2020 project NEXES and was extensively tested in a pilot setting at the University of Ljubljana, with a demonstration event with attendance of all major system stakeholders (first responders, PSAP operators, and solution vendors). The viability of the developed system was demonstrated in five complex scenarios, and extensibility of the PEMEA system was confirmed using multiple modes of communication and information sharing. Due to the simplicity of the underlying PEMEA protocol, multiple additional integrations have followed ours in a short time span, and the extensibility of the protocol could easily be leveraged by third-party communication apps to deliver a robust emergency calling experience. The next steps include further development and productization of the components to fit into the nascent PEMEA ecosystem supported by EENA, as well as tackling the usability challenges of the end users where steps need to be taken to raise awareness about emergency apps and make them more user-friendly and generally useful. Finally, we also want to address how the emergency systems like this can leverage telco-grade 5G network features, such as prioritized slices, QoS, direct mode communication, and fine-grained network positioning.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The research and development work was in part supported by the European Commission (Grant Agreement no. 653337), the Republic of Slovenia, and the European Regional Development Fund (project 5GSafety).

References

[1] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of WiFi, smartphone sensors and landmarks using the kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, 2015.

[2] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppänen, and M. Valkama, "High-Efficiency Device Positioning and Location-Aware Communications in Dense 5G Networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 188–195, 2017.

[3] M. Šterk and M. Praprotnik, "Improving emergency response logistics through advanced GIS," *Open Geospatial Data, Software and Standards*, vol. 2, no. 1, 2017.

[4] M. Ortiz, M. De Sousa, and V. Renaudin, "A new PDR navigation device for challenging urban environments," *Journal of Sensors*, vol. 2017, Article ID 4080479, 2017.

[5] Z. Deng, X. Liu, Z. Qu, C. Hou, and W. Si, "Robust heading estimation for indoor pedestrian navigation using unconstrained smartphones," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5607036, 11 pages, 2018.

[6] A. Stern and A. Kos, "Positioning performance assessment of geodetic, automotive, and smartphone gnss receivers in standardized road scenarios," *IEEE Access*, vol. 6, pp. 41410–41428, 2018.

[7] L. Vidmar, M. Stular, A. Kos, and M. Pogacnik, "An automatic Wi-Fi-based approach for extraction of user places and their context," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 154958, 15 pages, 2015.

[8] "ETSI TS 103 478 V1.1.1 (2018-03): Emergency Communications (EMTEL); Pan-European Mobile Emergency Application," Article ID 103400, https://www.etsi.org/deliver/etsi_ts/103400_103499/103478/01.01.01_60/ts_103478v010101p.pdf.

[9] F. Liberal, J. O. Fajardo, C. Lumbreras, and W. Kampichler, "European NG112 Crossroads: Toward a new emergency communications framework," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 132–138, 2017.

[10] EENA, "Term Definition Document," 2012, https://web.archive.org/web/20180115155320.

[11] EMYNOS, "nExt generation eMergencY commuNicatiOnS H2020 project," 2018, https://www.emynos.eu/.

[12] Y. Rebahi, K. T. Chiu, N. Tcholtchev, S. Hohberg, E. Pallis, and E. Markakis, "Towards a next generation 112 testbed: The EMYNOS ESInet," *International Journal of Critical Infrastructure Protection*, vol. 22, pp. 39–50, 2018.

[13] E. K. Markakis, A. Lykourgiotis, I. Politis, A. Dagiuklas, Y. Rebahi, and E. Pallis, "EMYNOS: Next generation emergency communication," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 139–145, 2017.

[14] NEXES, "NEXt generation Emergency Services H2020 project," 2018, https://nexes.eu/.

[15] D. Corral-De-Witt, E. V. Carrera, J. A. Matamoros-Vargas, S. Munoz-Romero, J. L. Rojo-Álvarez, and K. Tepe, "From E-911 to NG-911: Overview and Challenges in Ecuador," *IEEE Access*, vol. 6, pp. 42578–42591, 2018.

[16] I. Osebor, S. Misra, N. Omoregbe, A. Adewumi, and L. Fernandez-Sanz, "Experimental simulation-based performance evaluation of an sms-based emergency geolocation notification system," *Journal of Healthcare Engineering*, vol. 2017, Article ID 7695045, 12 pages, 2017.

[17] R. Öörni and A. Goulart, "In-Vehicle Emergency Call Services: ECall and beyond," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 159–165, 2017.

[18] R. Öörni and T. O. Korhonen, "ECall minimum set of data transmission - Results from a field test in Finland," *IET Intelligent Transport Systems*, vol. 8, no. 8, pp. 639–647, 2014.

[19] M. Cabo, F. Fernandes, T. Pereira, B. Fonseca, and H. Paredes, "Universal access to eCall system," *Procedia Computer Science*, vol. 27, pp. 104–112, 2014.

[20] A. H. Tapia, N. A. Giacobe, P. J. Soule, and N. J. LaLone, "Texting for large-scale disasters: developing practical technical innovations for emergency management at public universities," *International Journal of Public Administration in the Digital Age (IJPADA)*, vol. 3, no. 3, pp. 73–85, 2016.

[21] "Advanced Mobile Location (AML) Specifications & Requirements," 2016, http://www.eena.org/download.asp?item_id=165.

[22] "Pan-European Mobile Emergency Application (PEMEA) Protocol and Procedures Specification," 2018, http://www.eena.org/download.asp?item_id=175.

[23] J. Winterbottom, B. Casse, C. Lumbreras, P. Sanders, I. Gomez, and P. Kolios, "Pan-European Mobile Emergency Application (PEMEA) Requirements and Functional Architecture," 2018, https://eena.org/wp-content/uploads/2018/11/Pan-European-Mobile-Emergency-Application-PEMEA-Requirements-and-Functional-Architecture.pdf.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7802843, 19 pages
https://doi.org/10.1155/2019/7802843

Research Article

Resource Scheduling for Postdisaster Management in IoT Environment

J. Sathish Kumar and Mukesh A. Zaveri

Computer Engineering Department, SVNIT, Surat, India

Correspondence should be addressed to J. Sathish Kumar; sathish613@gmail.com

Received 27 November 2018; Revised 4 February 2019; Accepted 18 February 2019; Published 20 March 2019

Guest Editor: Maurizio Casoni

For postdisaster management, rescue and recovery operations are very critical. It is desired that the rescue and recovery operation should be handled through efficient resource management to minimize the postdisaster effects in terms of human loss and other types of damage. Resource management requires addressing various challenging issues like scheduling and monitoring of the resources which need real time information of various activities or events occurring anytime, anywhere, and anyplace. To satisfy such requirements, Internet of Things, an advanced upcoming technology, can be utilized for resource monitoring and scheduling. In this context, we propose resource scheduling algorithm for the postdisaster management. As mentioned above various tasks of rescue and recovery operation should be carried out with different priority and there should not be deadlock while availing the resources. In our approach, we estimate the waiting time using queuing theory for the availability of the resources for different activities that are to be performed at various locations. The simulation results of the proposed method are analyzed using different standard parameters like resource utilization and the waiting time for different activities. The proposed method is further visualized through real time annotation of resources and activities represented with the help of Google maps using android based application on the smartphone. The proposed algorithm is further compared in terms of computational complexity and fairness analysis for the effective utilization of the available resources.

1. Introduction

Natural disasters such as earthquakes, floods, hurricanes, volcanic eruptions, and tsunamis strike at various places on the earth every year with a great loss of human beings and the economy. United Nations International Strategy for Disaster Reduction (UNISDR) survey [1] reported that the number of disasters across the earth has been 346 during the year 2015. As a consequence of these disasters, 22,773 people lost their lives, 98.6 million people were affected, and an economic loss of USD 66.5 billion had occurred. In India alone, 38 million people were affected with different types of disasters and USD 29 million of economic damage occurred [2] during 2011-2015. From this information, it is clear that handling the disaster situation is very critical. So, there is a need for an efficient disaster management.

Among all the various phases in the disaster [3, 4], the response and recovery phases are crucial for postdisaster management. For quick response and recovery, there is a need for different resources and these resources must be distributed in such a way that various activities can be performed efficiently. This leads to a need for proper resource utilization as the number of resources is limited. In this context, scheduling of resources is required for completion of different activities or tasks for postdisaster management. Scheduling of resources in the response phase is crucial because it has to perform various activities like the evacuation of the people and shifting them to a safe location on an urgent basis. To handle disaster-related rescue operations, the real time need of various activities should be identified properly. For instance, in the context of disaster, various activities like gathering the information across the disaster places for establishing communication network are essential and critical. If a communication network is IP enabled and if it is possible to connect various resources using this IP enabled network, then it is possible to have information about the need of resources and whereabouts of various resources in real time. This information can be utilized for the distribution

and scheduling of resources for different activities. Such IP enabled network, having resources with sensor devices and RFID tag, is possible to realize using the Internet of Things (IoT). IoT technology is very influential that works based on Internet [5–7] and can help in a great extent to meet the requirements of the postdisaster management. In this paper, we propose algorithm for such IoT based networks. As mentioned above with the help of IoT, it is possible to access the information anytime, anyplace, and anywhere which helps in decision making.

Depending on the type of activity and its impact, various activities should be executed in a timely and proper manner. In such scenarios, it is necessary that each activity should be assigned the priority. Based on the priority of various activities, they should be given more importance while scheduling different resources. For example, the hospitalization of the injured people is having more priority compared to the reallocation of properties or handling the economic loss related activities. Further, the resources are scheduled so that all activities should be addressed with minimum waiting time for completing the respective activities. For estimating the waiting time of the various activities, queuing theory is explored. Oftentimes, disaster activities might take place in an overlapped manner with respect to time. This situation leads to a set of requests simultaneously for many resources by different activities at various locations which results in deadlock or race conditions. If rescue operations are not carried out at right time, the situation may deteriorate. From the above discussion, it is clear that efficient resource scheduling for various activities is necessary and challenging. In this context, we propose the priority based resource scheduling algorithm in this paper.

The main contribution and novelty of our paper is that we propose a resource scheduling algorithm in IoT environment for postdisaster management, by considering the priorities of the activities and to avoid deadlock conditions, with optimum number of resource centers by estimating minimum waiting time using queuing theory. For evaluating our proposed method, a case study is conducted in Surat city of India, by emulating the affected area, for considering different activities to be completed with the given number of resources. The algorithm output is visualized on android based application using Google map. The resource monitoring at various activities can be carried out effectively. The proposed queuing model is evaluated with respect to resource utilization and waiting time with the different number of resource centers by varying the arrival rate and service rate of activities. The proposed approaches are compared with the standard set of parameters like fairness in the allocation of resources for utility and computational complexity.

The rest of the paper is organized as follows. Related work is presented in Section 2. Section 3 introduces the problem formulation. The system model based on queuing theory for evaluating the waiting time for resource scheduling is detailed in Section 4. The proposed resource scheduling algorithm is described in Section 5 and in Section 6, the simulation results are presented. Section 7 concludes the paper with references at the end.

2. Related Work

In this section, the review of overall postdisaster management resource scheduling and allocation of various approaches is explored. A study conducted by Fraunhofer Gesellschaft on emergency and disaster management systems [8, 9] concludes that, besides the information and communication management, the other areas which are critical in disaster management task are viz., visualization, decision support, optimization and simulation, geographic information systems (GIS), simulation, and training. Two important activities in disaster situations are logistics support and evacuation. Various approaches related to evacuation have been reported in the literature [10, 11]. The approaches related to the logistics have been explored in [10, 12–15]. A generic approach to handle the disaster management has been proposed in [16, 17] using greedy and game theoretic based models. Regarding the resource allocation and scheduling during disaster situations, various approaches have been proposed in [18–21]

Optimal resource allocation in emergency management was carried out with the help of agent-based system using cost allocation by creating events with a severity of the disasters at various levels and verified with Pareto optimality test in [18]. This approach is not suitable for handling resource allocation for concurrent emergency events, in which the occurrence time of different events could overlap with each other. Resource allocation model with two stages with vulnerability analysis has been proposed in [20]. The approach proposed in [20] works at rescue team level. The higher authorities may not be able to coordinate the tasks efficiently due to lack of the information. For postdisaster management, centralized and distributed coordination is required. But for disaster management, as the whole task is very complex, the top level decision must be incorporated for designing the response strategy which was lacking in [20]. Dynamic scheduling based optimized resource allocation has been proposed in [21] using the genetic algorithm. The approach in [21] may not be suitable for varying the situations and for real time resource mapping as various critical factors are not considered and further, it lacks in terms of integration of various modules required for disaster management. In [22], scheduling using priorities transportation agency in hurricane recovery has been proposed.

For disasters like earthquake, the resource allocation approach has been optimized for emergency response using dynamic combinatorial optimization model [23]. The approach in [23], specifically proposed to handle earthquake situation, might not be suitable for other kinds of disasters. Distribution in humanitarian relief using mixed integer linear programming approach has been proposed in [24] using different parameters like transportation cost and a number of days needed to rescue. The approach reported in [24] may increase computational complexity if an optimized solution is required for a given situation. Allocation and scheduling of rescue units in natural disaster management has been proposed in [25] using Monte-Carlo based heuristic approach. As stated in [25], the performance could be improved by coordinating and collaborating over the autonomous agents.

For flood-related disasters, an approach for logistics preparation in an uncertain environment has been proposed in [26]. Similarly, for urgent relief in disaster situation based on demand, for logistics distribution has been proposed in [27].

In the literature, for handling emergency response scenarios the queuing theory has been used to address various issues. In [28], a detailed survey has been carried out for disaster management operations. Queuing theory has been explored for health care management services in [29–34]. Queuing theory based model [29] has been proposed for the patients requiring an immediate emergency cardiac response. Likewise, in [30], using queuing theory method, approach has been proposed for enhancing hospital emergency service. These approaches lack the integration of various processes and the automation of the complete system. In [31], a technique has been suggested to analyze the completion time for responding in case of accidents and emergency situations, which can be utilized by the respective departments of the UK government. In [31], it is stated that as the arrival rate of patients increases, then the system becomes unstable in estimating the targets of such accidents. Similarly, the allocation of various resources in ICU (intensive care unit) of the hospitals during busy times of emergency has been explored in [32]. As per the study carried out in [32], the allocation of resources in disaster situations yielded uncertain results due to randomness and unpredictability. Kristin Fitzgerald et al. [33] proposed a queue-based Monte-Carlo analysis in disaster situations for decision making to execute the rescue operations by the hospital management as fast as possible. In [33], the waiting time for the patients to get the nursing resource as demanded was evaluated. In the approach proposed in [33], worst case scenarios have not been considered. For disaster circumstances, the solution to handle worst case scenarios is very much crucial and needed at the highest priority. Likewise, in [34], the queuing theory has been exploited to make the analysis for managing the hospital systems during emergency times efficiently. The waiting time for the patients for availing the treatment was not discussed in [34].

The approaches that have been discussed so far give a glimpse of the resource scheduling and allocation in disaster management situations, but to understand the current scenarios, we explored various approaches that have been recently proposed [35–40]. The resource allocation approach for postdisaster management in the context of IoT has been proposed by considering the priorities of both various activities and resources using stable matching approach [35]. Likewise, resource allocation for maximum utilization of the resources using maximum bipartite graph theory has been proposed in the context of IoT for postdisaster management [36]. In both approaches of [35, 36], the estimating of the waiting time for allocation and number of resource centers to be utilized has not been detailed. The approach to optimize the decisions related to logistics in a collaborative way for the flood disaster situation has been proposed [37]. But, it is reported in that the proposed model is not able to manage the differences between the various organizations involved for rescue and their utilization to an essential required place [37]. Likewise, an approach for postdisaster humanitarian logistics in the context of proving the medical assistance teams has been proposed [38]. However, few limitations are reported that the classification of various types of relief medical supplies is not carried out and while dispatching the logistics, heterogeneous vehicle routing problem is not addressed. Further, resource scheduling approach is proposed in the IoT environment for postdisaster management [39]. But, the drawback in this approach [39] is that the estimating of the waiting time for various tasks and resources is not addressed. An approach for allocation of the resources during simultaneous disasters has been proposed [40]. They performed the risk analysis with resource constraints for meeting the given targets and estimated the financial aspects with different budgets and further the political effects have been considered.

From the above literature, it is clear that there is a scope for an efficient resource scheduling algorithm for postdisaster management. In this context, we propose an algorithm which considers the different critical parameters for allocation and scheduling of the resources in an effective manner. The proposed scheduling algorithm along with the problem formulation is described in the following sections.

3. Problem Formulation

For postdisaster management, the rescue and recovery operations are very crucial and to carry out these operations resources are required. Having limited number of resources, for various task accomplishment, resource scheduling must be done in an efficient manner. Scheduling of the resources can be performed using the number of parameters. In this paper, we consider two parameters, namely, the priority of the activities and the expected completion time of an activity. The completion time of any activity depends upon the resource availability. In such case, the resource availability is very crucial and must be handled efficiently. In this context, it is very important to find out time duration after which the resource may be available for a given task or activity. This time duration can be considered as a waiting time. So, it is necessary to estimate a waiting time for resources and it indirectly provides a measure for the completion time of an activity. In our proposed algorithm, we estimate the waiting time for the availability of the resources that is required for a specific activity. Generally, various resources are dispatched from different locations in the affected region. These locations are considered as resource centers. At the beginning of any rescue operation, various resources are distributed from different resource centers. It is very important to decide the number of resource centers that should be established to cover the maximum affected area. Similarly, the disaster affected area should be divided into subareas or zones for effective rescue operations. Each zone may have at least one resource center. In this context, the required optimal number of resource centers needs to be estimated for the postdisaster management activities.

In brief, from the above discussion, the waiting time for the availability of a resource and the number of resource centers required to carry out rescue operation are significant

and this operation should be performed in a very efficient and effective manner. In this view, we exploit a very standard approach, queuing theory, for estimating a waiting time and the number of resources and using these estimated parameters, the resource scheduling algorithm is introduced. It is also required that the various activities should be completed in a particular order, which leads to a situation where the given activities should be completed with a particular priority which may be preassigned. The waiting time evaluation and estimating the optimal number of resource centers and the priority assigned to different activities are the key parameters for effective and efficient postdisaster management. Further, the scheduling of the resources must be done in a fair manner avoiding the race conditions or starving among the activities. For scheduling, a large number of methods have been reported in the literature which is detailed in the related work section. Among various scheduling algorithms surveyed, we explore Banker's algorithm [41, 42] for our proposed scheduling algorithm. Our proposed algorithm is enhanced in such a way that it should meet the required conditions like the priority of a given task or activity. In the literature, Banker's algorithm has been used for process scheduling in the operating system. Earlier, we have proposed a scheduling approach [43] for postdisaster management using Banker's algorithm. In this paper, our earlier proposed algorithm is enhanced by incorporating the parameters, namely, the priority of the task and waiting time for the availability of the resources for resource scheduling.

From the above discussion, for our proposed system, we address mainly three issues which are very important and crucial for rescue and response operations as a part of postdisaster management. These three issues are the number of resource centers to be utilized and resource scheduling of the respective activities by considering the priorities and scheduling them without deadlock conditions. For this purpose, the disaster affected region may be divided into regions or zones for effective activity management. All the available resources are to be visualized at the top level as a group. In this context, the problem formulation is represented in the form of a graph model with various resources and their interconnections in the form of activities.

Let us say that there are l activities and m resources in the system. These activities say $a_1, a_2, \ldots, a_l$ can be denoted by a_i, where $(1 \leq i \leq l)$. Likewise, the resource of j type is termed as r_j, where $(1 \leq j \leq m)$. In the system, it is assumed that the number of similar types of resources is available, i.e., if a particular resource r_j is available with k number of quantities, then we define it as k instances of r_j. The number of instances of k varies from one resource to other resources. This information is available at the time of planning. For example, three ambulances are present in a system considered as three instances of resource type ambulance.

In a graph G, each node is labeled, either, as an activity node or a resource node. The node is represented as an activity node, a_i, and resource node, r_j. These particular activity nodes and resource nodes are interconnected through the edges. In real practice, the rescue and response operation have to perform the number of tasks. Each task consists of different activities which results in dependencies among the activities for completing a task. In our proposed system, the dependency among the activities is handled at the scheduling level. In this view, the typical graph using activities and resources looks like as shown in Figure 1. To make graph representation simpler, the different instances of resources are also represented as graph nodes and these instances are represented by black dots within the resource node itself. It is assumed in this example that there is no activity dependency. So, the activity nodes are not interconnected. Figure 1 reflects the same, where r_1 and r_3 have one instance each, whereas r_2 and r_4 have two and three instances, respectively.

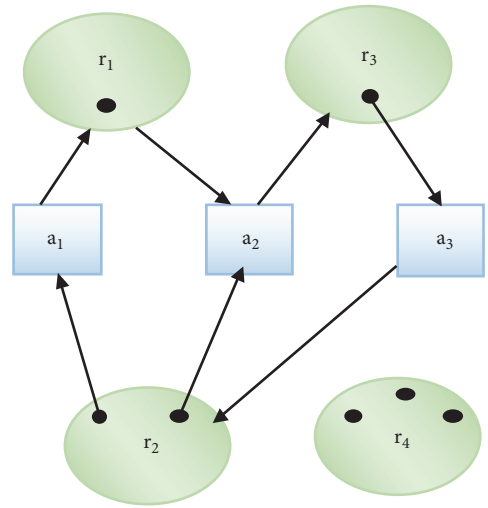


Figure 1: Illustration of improper scheduling.

As pointed out earlier, various tasks consist of different activities and each activity requires a set of resources. It may result in deadlock or race condition among the different activities. As mentioned, the dependency among the activities is handled by the scheduling algorithm but it is necessary to map these dependencies in the graph of activities and resources nodes. The scheduling algorithm is devised in such a way that it should help in resolving the deadlock or race conditions. In such cases, it may happen that different activities require the same set of resources for completing the task. As each activity needs a set of resources that need to be scheduled, the allocation of resources to the activities and scheduling them must be handled simultaneously to avoid the deadlock or race situations. It is important to note that, in our proposed systems, we take care of this critical requirement through the graph which helps the scheduling algorithm for fair allocation. In this view, the resource dependencies among the activities are represented through directional edges in the graph. The arrow from an activity node to the resource node indicates that a particular activity needs that resource. The arrow from resource node instance to activity node indicates that a particular instance of the resource is allocated to that activity. Indirectly, this representation of activities and resources allows one to find a cycle or deadlock situations or race conditions at a given instance of time. All these directional edges represent the demand in the allocation at a given instance of time and that demand varies with the time.

The major contribution of this work is to represent the demand and the allocation of the resources and dependencies among the activities, through the directional graph. This makes our graph dynamic which actually represents a dynamic network of resources. This graph representation again helps the management authority at a given instance of time for monitoring the resources and activities and hence, the task at any level can be performed with ease. As shown in Figure 1, activity a_1 needs resource r_1 which is currently allocated to activity a_2. Similarly, two instances of r_2 are allocated to activities a_1 and a_2, respectively. But the same instance resource r_2 is required by the activity a_3. As there is no additional instance of r_2 available, it cannot be allotted to activity a_3. This requires an effective and efficient scheduling. The demand and allocation are represented as a directed graph $G = (V, E)$, where V is a set of vertices that indicates the activities and resources and $E \subseteq \{\{a_i, r_j\} \mid a_i, r_j \in V, a_i \neq r_j\}$ defines the potential allocation edges. Any resource scheduling algorithm should perform the scheduling of resources among the activities in a fair manner. There should not be any deadlock situation or race condition which results in the starvation of resources which holds the complete system in a standstill condition. So, it is important to schedule the different activities and allocate the resources based on some additional information or the number of parameters. These parameters should be chosen in such a way that it should lead to a very effective and efficient resource allocation and scheduling for postdisaster management activities.

These parameters are utilization of the optimum number of resource centers, evaluating the waiting time for availability of the resources for completing a particular task, i.e., waiting time for a particular resource, and resource scheduling discussed earlier with the priority of different activities. In this view, the system must have a methodology to incorporate these three parameters in resource allocation and scheduling. The postdisaster management requires real time dynamic decision making. For supporting the dynamic decision making using these three parameters, it is necessary to evaluate them in real time, i.e., based on the demand of the allocations. Considering this fact, the best solution that one has is based on queuing theory. In this context, our system is modeled using queuing theory for evaluating these parameters. In the following section, we describe the queuing based modeling using the above parameters.

4. Queuing Theory Based System Model

In the literature, the queuing theory has been very well exploited for the problem of demand and supply or producer and consumer based applications. For demand and supply based applications, the system has the number of service providers and the number of service users. In queuing theory, various parameters like arrival rate of service user and service rate, length of the requests in the queue, and waiting time of the requests in the queue are very critical and modeled based on statistical distribution. In the context of postdisaster management, the task and activities are the services to the user or consumer, whereas resources act as a service provider or producer. In this view, the queuing theory based modeling helps in designing the robust system which not only provides an optimal solution but allows one to complete the required task in an effective and efficient manner. The performance of the queuing based system depends on the number of available resources, i.e., service centers or servers in the system and the length of the requests in the queue. Consequently, these two parameters decide the waiting time for the availability of a resource for a particular activity. As discussed earlier, these resources are distributed and located at various places in the disaster affected area.

The queuing system is represented using Kendall's notation [28]. The Kendall's notation consists of four variables that define the system parameters. The standard Kendall's notation is given as M/M/C/K; then these parameters are defined as arrival rate/service rate/number of servers/capacity of each queue. For instance, if the arrival rate is represented as λ/hour and the service rate as μ/hour, with 1 server of the capacity of 10 items in the queue being hold, then the Kendall's notation can be given as //1/10. In the context of the disaster management, the various activities are requesting the resources to accomplish the given job. Hence, the resources are acting as servers and activities are waiting in the queue to get served by the resources. These resources are available in various resource centers. Therefore, the queuing theory in our system is defined as arrival rate of requests by various activities/service rate of the resources/number of resource centers/number of activities waiting in the queue.

Generally, the number of requests is not known in advance. In such a case, the queuing system can be modeled using an infinite number of requests. Similarly, for serving these requests, the resources available in the system are considered to be finite. The arrival rate and the service rate are modeled based on different probabilistic distributions. For instance as shown in the above notation, $M/M/C/K$ represents the modeling of arrival rate as Poisson distribution M as the first parameter, followed by the second parameter for service rate which is modeled as exponential distribution M. C is the number of resources centers which are available in the system and K is the number of requests arriving to the system by various activities. The reason for using Poisson distribution for arrival rate is because it involves memoryless waiting time until the arrival of the next request in nonoverlapping time intervals that are probabilistically independent.

For the postdisaster management, we explore the queuing theory and model the system accordingly. Postdisaster management consists of various tasks to be accomplished. These tasks are decomposed into different activities and each activity requires a set of resources for its completion. For our system, an arrival rate of the requests is nothing but the arrival of activities for different resources. The request is originated by a particular activity for a given task. Similarly, a service rate of different requests is nothing but a particular resource utilized by a particular activity of a given task. As discussed in the problem formulation, the service rate, i.e., the time for which a resource is utilized by an activity affects or decides when the same resource will be available to another

activity, which is considered as a waiting time for the next activity to acquire that same resource. In a queuing theory, this waiting time is also modeled similarly to the service rate; i.e., if the service rate is modeled using exponential distribution, then waiting time also follows an exponential distribution. The main aim of our work is to distribute the resources to different activities with minimum waiting time. In this context, we model our system using queuing theory to estimate the optimal number of resource centers required to minimize the overall waiting time for all the activities in the system. These resource centers can be utilized for efficient resource scheduling. The queuing system based modeling is detailed as below.

In this paper, for modeling the system using queuing theory, we define four parameters. These parameters are related to (i) the number of requests that arrive for different resources at resource centers, (ii) the service rate or execution time for completing the requests, (iii) the length of the queue for holding these requests for different resources at resource centers, and (iv) the capacity of the systems which is a measure of how many requests can be handled by the system at a given time, i.e., the number of the requests in a queue and the number of requests in a service or execution and overall time from entry to exit of a request through the system. Here, the system consists of the number of resource centers and each resource center has a number of resources. In this context, these four parameters are, namely, categorized as (i) the length of the activities in the queue to hold the number of requests that are coming to the system, (ii) the waiting time for the activities in the queue as the resources and the resource centers are busy with serving the other activities, (iii) the length of the total activities which is the capacity of the system holding the number of requests in the queue as well as in the service, and (iv) the waiting time of the total number of the activities which reflects the duration from entry to exit through the system. Determining the optimal number of resource centers and the waiting time for the requests using the queuing model based on the above four parameters is detailed as below. Now, the total number of activities in the system is defined in

$$\text{Total number of activities in the system} = \frac{\text{Activities arrival rate/Resources service rate}}{1 - \text{Activities arrival rate/Resources service rate}} \tag{1}$$

Since the length of the system, total number of activities in the system, is known, the number of activities that are waiting in the queue, waiting time of the all the activities, and waiting time in the queue can be determined using Little's equation [28] as shown in

$$\begin{aligned}\text{Total number of activities in the system} = {} & \text{Total number of activities in the queue} \\ & + \text{Expected number of activities are being served by the resources}\end{aligned} \tag{2}$$

$$\text{Waiting time in the queue} = \frac{\text{Total number of activities in the queue}}{\text{Activities arrival rate}} \tag{3}$$

$$\text{Waiting time of all activties} = \frac{\text{Total number of activities}}{\text{Activities arrival rate}} \tag{4}$$

The example of the single resource center is explained with an example in Appendix A.

The above model is derived using a single resource center. The waiting time for the availability of resources for different activities which are served by a single resource helps further to design the system with multiple resource centers. In real practice, there are many areas in the affected region where the different tasks are to be performed and only one resource center is not sufficient. Using the above model which is derived for a single resource center can be extended for multiple resource centers so that the waiting time for the availability of different resources for different activities of various tasks can be minimized and the rescue and recovery operation can be executed efficiently. In this context, the queuing system is modeled with number of resource centers that can be determined by the following equation

$$\text{Total Number of Activities in the queue} = \frac{\text{Activities arrival rate/Resources service rate}^{\text{Number of Resource Centers}+1} P_0}{(\text{Number of Resource Centers} - 1)!\,(\text{Number of Resource Centers} - \text{Activities arrival rate/Resources service rate})^2} \tag{5}$$

where, P_0 is the probability that there is no queue. Equation (5) represents the relationship between the length of the requests in the queue and the number of resource centers. The optimal number of resource centers can be evaluated by verifying by (5) =1, i.e., no waiting time. It means that as soon as the requests come, they will be processed which depicts an

```
(1) procedure Scheduling(Available_Curr,
    Allocation_Curr, Demand_Curr, Requested, Priority_p)
(2)     Schedule the activities by sorting from high to low priority
(3)     while for all activities w.r.t Priority_p do
(4)       if Requested ≥ Demand_Curr then
(5)         Go for next priority activity          ▷ Activity exceeded the max claim
(6)       else if Requested ≥ Available_Curr then
(7)         Go for next priority activity          ▷ Due unavailability of resources
(8)       else                                     ▷ Ensuring the availability and priority
(9)           Available_Next = Available_Curr - Requested   ▷ Allocating the requesting resources
(10)          Allocation_Next=Allocation_Curr+Requested     ▷ Decrement the demand after allocation
(11)          Demand_Next = Demand_Curr − Requested
(12)      end if                  ▷ Update the Available resources
(13)      Available_Curr = Available_Next
(14)    end while
(15)    Return schedule and execute the activities.
(16) end procedure
```

Algorithm 1: Resource Scheduling Algorithm (Without Waiting Time Estimation).

ideal situation. The example of the single resource center is explained with an example in Appendix B.

5. Proposed Resource Scheduling Algorithm

In the postdisaster management, rescue and response operations are very crucial. The rescue operations should be carried out effectively in such a way that various services should be restored and need of the affected people should be addressed. For this purpose, acquiring the data and availability of this data anytime, anyplace, and anywhere can be achieved through IoT.

Further, different IoT devices, sensing various events like fire, temperature, and gas leakage, detectors of human life, etc., are deployed at the affected area after the disaster. These devices may be useful for tracking the further stroke of such disasters because IoT devices help in managing and tracking the different resources in the affected area. In such scenarios, it is necessary to group all these devices so that the information acquired through devices can be channelized or utilized for rescue work. In this context, it is important to localize these devices which are deployed through drone like equipment [44]. It is assumed that localization of all these devices is carried out using [45]. Similarly, for effective data acquisition and information extraction, all these devices must be clustered. For clustering, our earlier proposed approaches [46–49] are used.

In a real application, the number of resource centers is limited and constrained by the number of other issues. So, it is difficult to achieve the state of the system, where there is no waiting time. It means that the number of requests is more than the number of servers/resources. In this scenario, it is required to maintain the queue and schedule the activities in such a way that there are no starvation and race conditions. In the literature, various approaches have been reported for resource scheduling as discussed in Section 2. Using our system model described above, we present the scheduling algorithm as shown in Algorithms 1 and 2. Among various algorithms reported in the literature, Banker's algorithm [41, 42] has been proposed for resource scheduling for avoiding the deadlock situations. The limitation of this Banker algorithm is that it requires in-advance knowledge of the number of resources required which is not possible in real applications. But at the same time, the advantage of Banker's algorithm is that it avoids the deadlock conditions and schedules the resources in a safe sequence. For a particular disaster management, to initiate the rescue and recovery operations, the work in the affected area is generally started with the available number of resources. In this context, Banker's algorithm is more suitable for the resource scheduling in a disaster affected area. Earlier, Banker's algorithm has been explored for resource scheduling in [43]. In a real environment, the various tasks which are to be performed in a disaster affected area have different priorities so that there will be minimum human and financial loss. So it is important to incorporate the priority of the task in allocating the resources or scheduling the resources. Second, the number of resource centers will be limited and the number of requests will be more. In such a situation, for completing the different activities of a task, it has to wait for the availability of the resources from the resource centers which introduces the waiting time for the requests to be executed. This discussion shows that the priority of the activity and waiting time for the availability of the resource must be considered in scheduling the different resources. This motivates us to propose modified Banker's algorithm which incorporates the priority of the tasks and the waiting time for the availability of a particular resource for completing a particular activity of a task. Queuing theory based modeling described in the previous section provides the queue length for the resource and waiting time of the system. As discussed in the problem formulation, let us assume that there are l activities and m resources in the system. The activities are denoted by $a_1, a_2, \ldots, a_l$ are given as a_i, where $(1 \leq i \leq l)$ and, likewise, the r_j, where $(1 \leq j \leq m)$ with k instances that vary from one resource to another. The total number

(1) **procedure** Waiting Time Estimation(Arrival rate, Service rate and Number of Resource Centers)
(2) **while** for i in all activities a_i **do**
(3) Calculate the values as per equation (1) to (5). ▷ Estimate waiting time for activities
(4) Determine the optimal number of resource centers using (5)
(5) Verify the waiting time is ≈ 0 using multiple resource centers.
(6) **end while**
(7) Call Algorithm 1;
(8) **end procedure**

Algorithm 2: Resource Scheduling Algorithm (With Waiting Time Estimation using Queuing Theory).

of resources of type j required by an activity i over the time to complete the activity is represented as $Demand_{ij}$. Similarly, the number of resources of type j allocated to an activity a_i is indicated by $Allocation_{ij}$. This representation leads to the two-dimensional matrices demand and allocation represents the current state of the requests for a particular resource and number of resources in service. As the requests arrive for the different resources from the different activities, $Demand_{ij}$ varies. As the resources are allocated to a particular activity, $Allocation_{ij}$ is updated. As discussed earlier for a resource type r_j, the number of currently available instances is stored in $Avaliable_j$. The particular type of the resource may be required by different activities of different tasks. So it is not possible to allocate all the available resources for a particular activity at a given moment of time. The $Allocated_{ij}$ is the number of resources of type j currently allocated to activity a_i against its $Demand_{ij}$. So, additional resources of type j required by an activity a_i for its completion are stored in $Requested_{ij}$. With these definitions, now we are able to verify the resources that are available and can be allocated immediately. The first criterion checks the requested number of resources that should be less than the available number of resources; i.e., the additional required resources are less than the total available resources which can be represented as (6). The proposed algorithm fulfills the essential conditions of the activities and verifies the availability of the requested resources using the given equation

$$\sum_{i=1}^{l}\sum_{j=1}^{m} Requested_{ij} \leq \sum_{j=1}^{m} Available_j \tag{6}$$

for l number of activities, i.e., a_i where i=1 to l, and j number of different types of resources r_j where j varies from 1 to m. This criterion presents that a particular activity needs a different type of the resources for its completion; i.e., the equation indicates that the total number of additional resources required by the different activities is less than the total number of available different resources. Similarly, the second criterion states that the demand is more than the available resources; then the request is placed in the queue. This can be as shown in

$$\sum_{i=1}^{m} r_i \leq \sum_{i=1}^{l}\sum_{j=1}^{m} Demand_{ij} \tag{7}$$

for l number of activities, i.e., a_i where i=1 to l, and j number of different types of resources r_j where j varies from 1 to m.

With these formulations, the algorithm is represented with a simpler resource scheduling algorithm based on Banker's method with priority assigned to each activity. The flowchart of the proposed algorithm is shown in Figure 2 and corresponding pseudocode is depicted in *Resource Scheduling Algorithm* 1. In the proposed algorithm, the subscript *Curr* and *Next* are used with the different notations to represent the current value of that variable and the value available for the next iteration, respectively. For example, $Available_{Next} = Available_{Curr}$ - *Requested* represents the number of resources of a particular type available for the next iterations which can be evaluated by subtracting the additionally required number of resources by a particular activity from the currently available number of resources of a particular type.

In this context, two variants of proposed algorithm are presented for resource scheduling. The first variant considers only the priority assigned to the activity without considering evaluation of waiting time for availability of a resource. Second, we represent scheduling algorithm where the priority assigned to different activities and waiting time for availability of a particular resource is also considered which makes our resource scheduling algorithm dynamic and elegant. Algorithm 1 presents resource scheduling algorithm considering the priority of the task or activity on hand and Algorithm 2 estimates the waiting time for the resource availability for completing different activities using queuing theory and calls Algorithm 1. To evaluate the waiting time in the queue, parameters like arrival rate, service rate, and number of resource centers are important and essential. In our proposed approach, by keeping number of resource centers constant, we observe the waiting time by increasing the service rate for different arrival rates. The same information is presented as a flowchart in Figure 3 and corresponding pseudocode is shown in Algorithm 2. Upon the observation with different arrival rates, the optimum number of resource centers can be utilized. After the resource centers are finalized, the scheduling algorithm is called for the resource allocation for various activities.

The time complexity of the proposed *Resource Scheduling Algorithm* 1 is $O(a^2r)$ with utmost a possible activities and r resources, whereas the brute force approach time complexity is $O(a!/(a-r)!)$. For a number of activities and r number of resources, the enumeration in FCFS approach [50] is ${}^aP_r$ which is equal to $(a!/(a-r)!)$. Therefore, the proposed

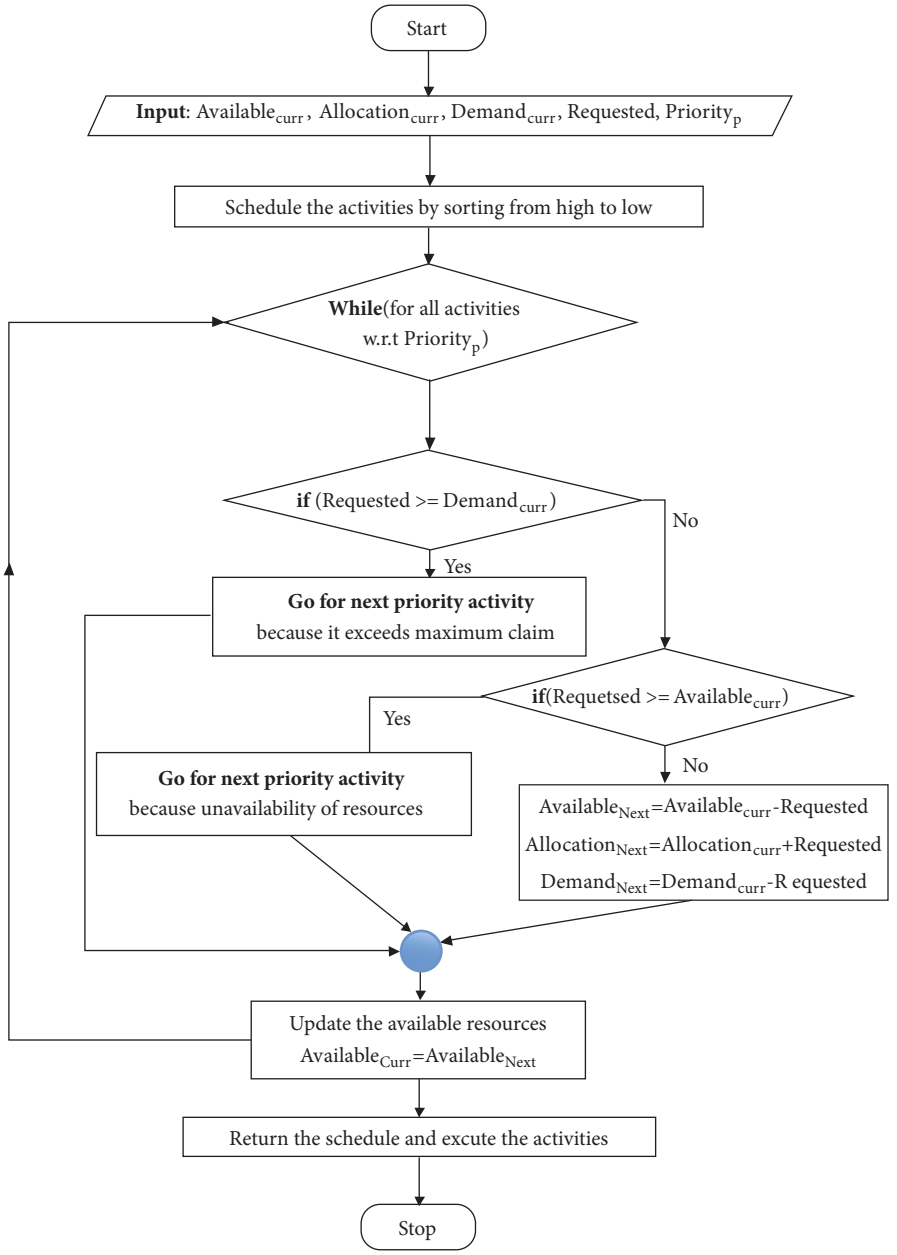


Figure 2: Flowchart of the Proposed Resource Scheduling Algorithm 1.

resource scheduling algorithm using Banker's strategy gives optimal results compared with the brute force approach.

6. Simulation Results

In this section, the simulation result is presented in three subsections. The first subsection presents the evaluation of queuing based proposed system whereas second subsection describes the result of evaluation of proposed priority based scheduling algorithm. The simulation results are carried out for evaluating the proposed queuing theory based system. The simulations are performed using different scenarios for evaluating the waiting time and determining the optimum number of resource centers given queuing system parameters like arrival rate and service rate. In each scenario, the aim is to minimize waiting time and complete different activities of various tasks required for rescue and recovery operation for postdisaster management. Efficient scheduling of various resources through different resource centers should be allocated in an optimal way. A case study is presented having different emergency locations in the area where different activities are to be performed using Google map. For this case study, the number of resource centers with different types of resources is considered. These resources are required by various activities of different tasks that are to be performed for rescue and recovery operations. The proposed scheduling

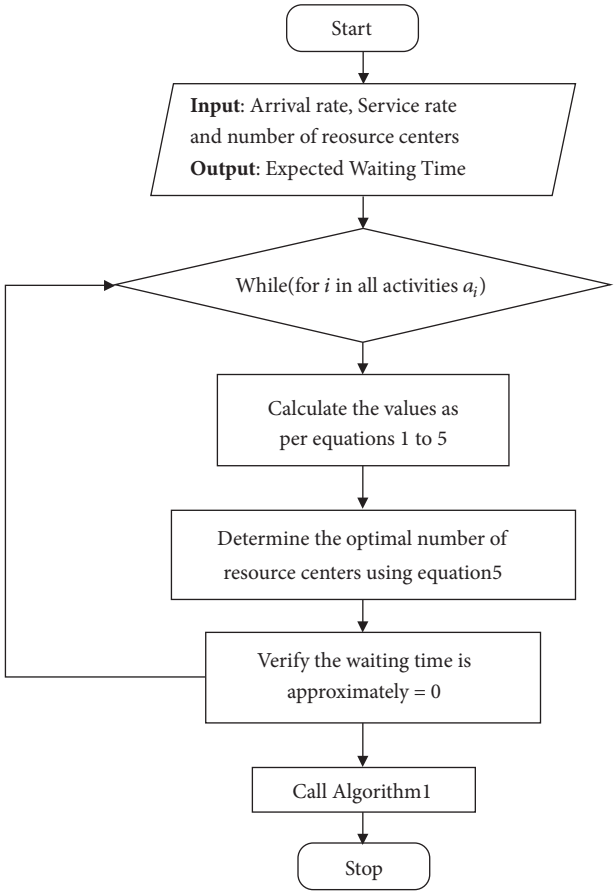


Figure 3: Flowchart of the Proposed Resource Scheduling Algorithm 2.

algorithm outcomes are depicted in the visual form using the developed application on the android platform which can be deployed for postdisaster management. For the actual implementation purpose, we designed and created our own android application using API:19 Android 4.4 (KitKat) [51]. In the third subsection, the comparative analysis is presented. The proposed algorithm is evaluated and compared with existing approaches reported in the literature. The performance analysis is carried out using different standard parameters, algorithm computational complexity time, and fairness in resource allocation and scheduling.

6.1. Evaluation of the Queuing Based System. For demonstrating the robustness of our queuing theory based system, the proposed model is evaluated in terms of resource utilization and waiting time with the different number of resource centers and activities. For ease of understanding, we used few notations to represent the various parameters, such as arrival rate of the activities= λ, service rate of the resources = μ, number of resource centers =C, and resource utilization=$\rho = \lambda/\mu$. With these notations, the simulations are performed in different scenarios. Initially, the resource utilization is evaluated by keeping the number of resource centers constant. For the given number of resource centers, the simulations are performed by varying the arrival rate of requests originating from different emergency locations for various activities. The resource utilization is plotted with respect to increasing rate of service for different requests that arrive at resource centers. It is shown in Figure 4.

It is clearly observed from Figure 4 that the resource utilization decreases as the service rate increases for different arrival rates. As the number of resource centers increases, the resource utilization decreases which can be visualized from Figure 4. As shown in Figure 4, as the resource center increases from 1, 2, 3, and 5, the maximum resource utilization decreases from 1.0, 0.5, 0.35, and 0.2. Similarly, another parameter that is evaluated is the waiting time. For the different arrival rates, the waiting time for availability of the resource for different activities is plotted with respect to increasing rate of service which is shown in Figure 5.

As the service rate increases, the waiting time decreases for different arrival rates. Further, the waiting time is decreasing as the number of resource centers is increasing. As shown in Figure 5, as the resource centers increase from 1, 2, and 3 the maximum waiting time decreases. The corresponding values for resource centers 1, 2, and 3 are 1.0 hrs, 0.025 hrs, and 0.003 hrs, respectively. Also, it is observed that if the number of resource centers is increased more than 3, the waiting time of the various activities is almost zero as the service rate increases. Based on the above experimental results, it is observed that as the number of resource centers increases, the waiting time for availability of the resource is decreasing. But at the same time, the cost of the capital amount for setting up the resource centers increases as the C increases. Hence, there should be a tradeoff in utilizing the optimum number of resource centers such that the waiting time for the activities can be handled effectively.

6.2. Evaluation of Proposed Resource Scheduling Method. For evaluating the proposed modified Banker algorithm for resource scheduling, the real scenario is emulated as a case study. The Google map is used to show different emergency locations in the area chosen for emulation. These are the areas of local city, Surat, of the state Gujarat in India. The complete visualization of resource scheduling is executed through the android application developed for smartphone devices. The idea is on-site application can be used in case of disaster and various resource centers can be monitored in real time. All devices which are part of resources may be IP enabled and may serve as IoT devices. These IoT devices now can be monitored through our developed application anywhere and anytime. This is the advantage of having IoT enabled environment for postdisaster management. All the tasks and various activities of these tasks along with their utilization of various resources can be monitored using a simple mobile device. The network can be accessed using a simple plugged router as an access point in the affected area. The proposed modified Banker algorithm is implemented on android platform for online monitoring of the complete network. The resource scheduling scenario is tested and analyzed. The various screenshots of execution of resource scheduling method are presented using Google map [52].

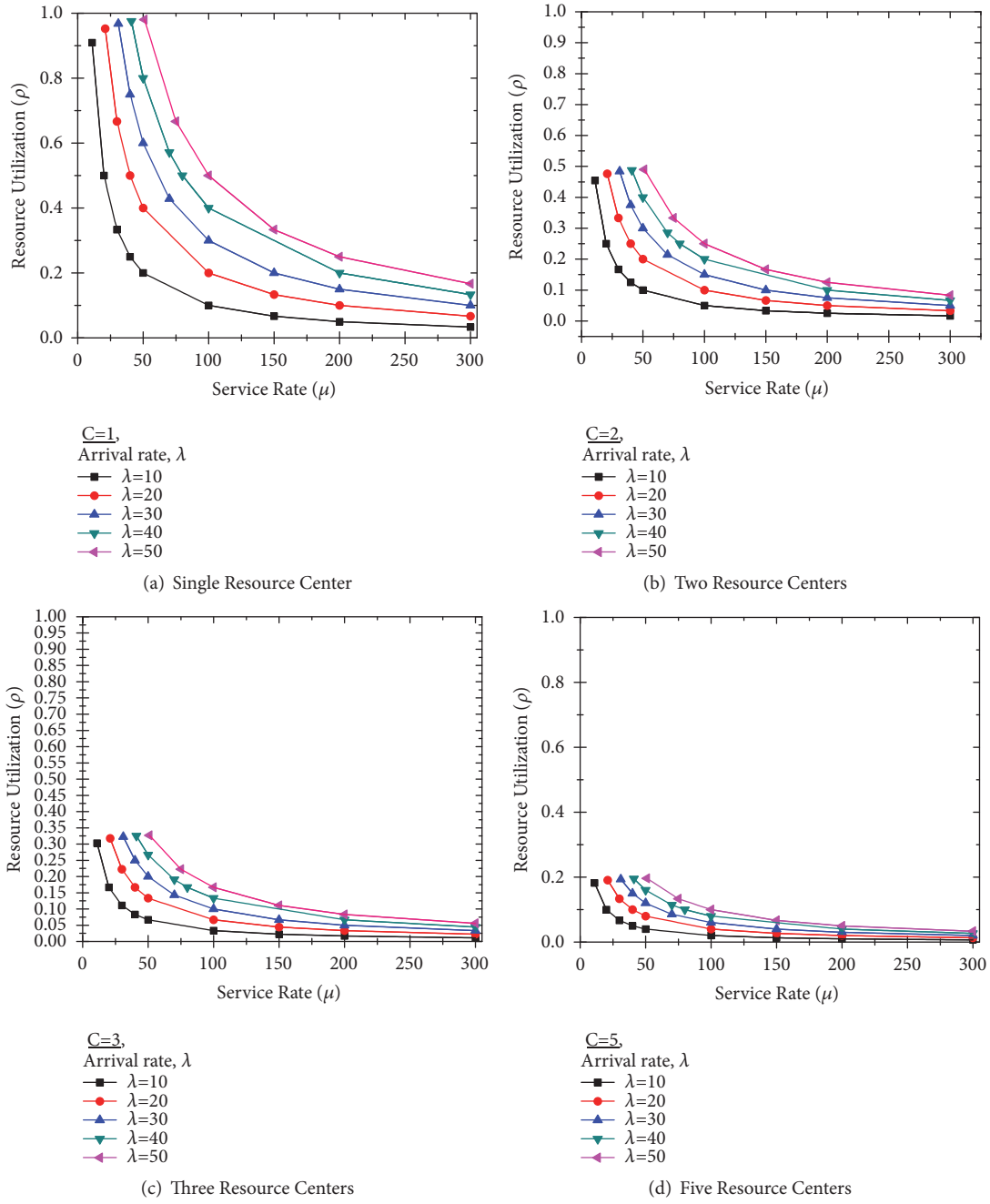


Figure 4: Resource Utilization of Various Resource Centers.

The resource scheduling is evaluated by considering different areas of Surat city as emergency locations. Let us assume that the resources are requested by different tasks at emergency locations that are local areas, namely, VR Mall, Big-bazaar, SVNIT, Adajan, and Athwa Gate. The resources are requested from the different resource centers located at different areas of the city, namely, Adajan, Majura gate, and Katargam. The locations of emergency places and resources centers are visualized using Google map. These locations are shown in Figure 6. Figure 6(a) represents the particular screenshot where the emergency locations and resource center places are listed. Figure 6(b) displays the locations of resource centers using Google map. Figure 6(c) shows the emergency locations in the Surat city.

For compact visualization on a small screen of the mobile device, the symbols are assigned as R1, R2, and R3 for resource centers. Similarly, the emergency places are assigned names A_0, A_1, A_2, A_3, and A_4. These A_0, A_1, A_2, A_3, and A_4 may represent the different tasks or activities at respective places or locations. This representation is shown in Figure 7. Various activities of different tasks to be performed at emergency locations require a different number of resources from the

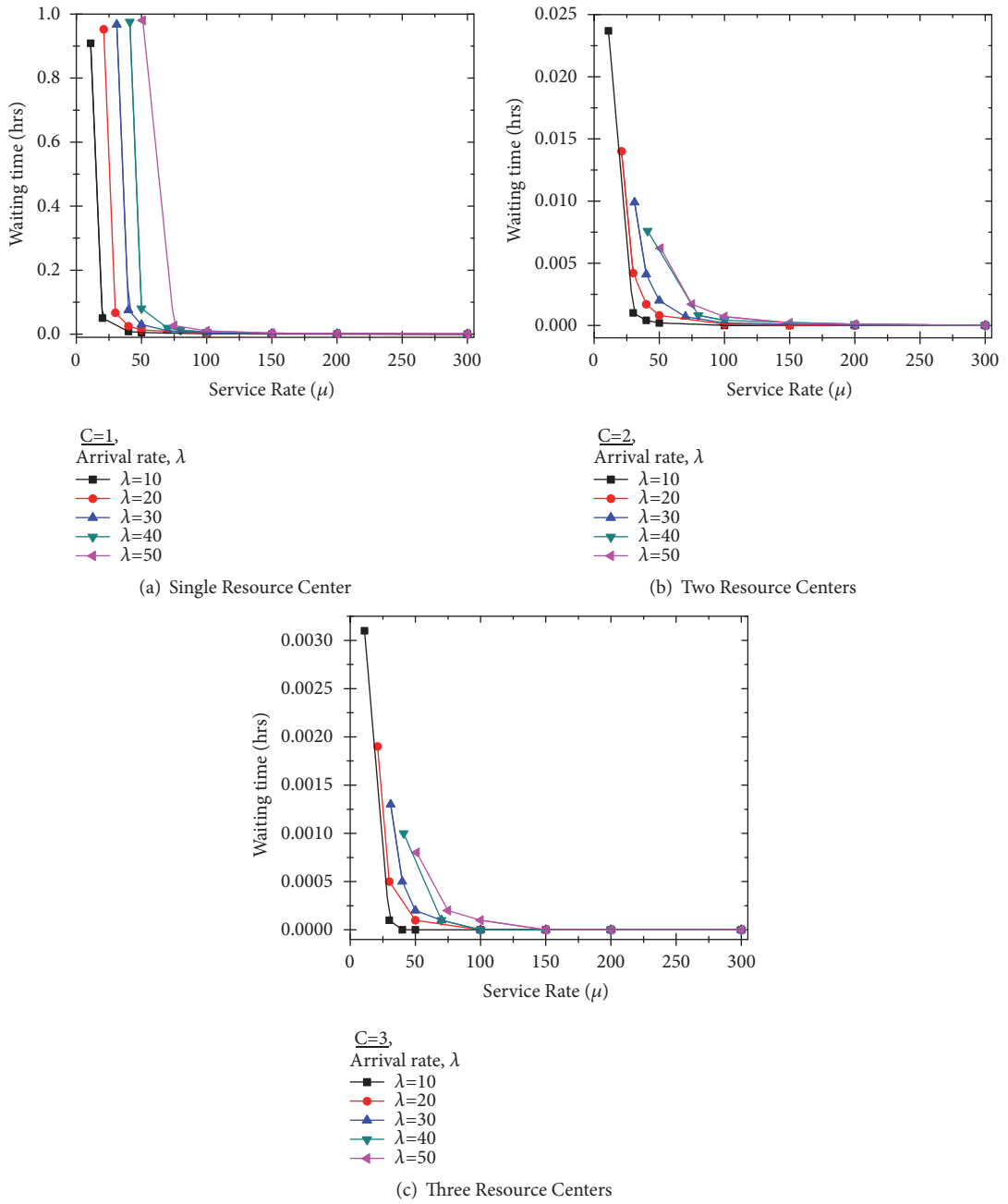
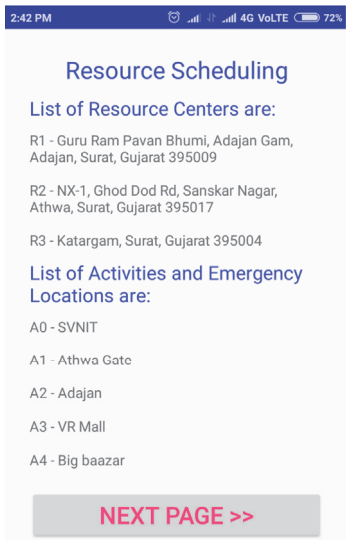


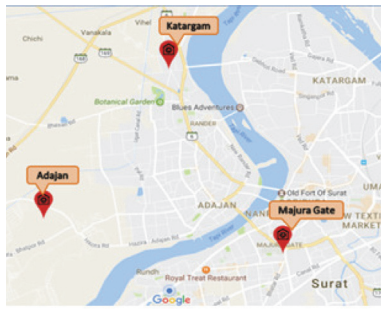
FIGURE 5: Waiting Time of Various Activities.

respective resource center which is represented as two-dimensional entries in the matrix form. As discussed in the proposed algorithm section, these two parameters are incorporated for better resource allocation and scheduling, i.e., the priority of the activity and the waiting time for the availability of the resource for an activity. To minimize the waiting time for the availability of the resources for different activities, we utilized queuing based modeling which is incorporated in the proposed resource scheduling algorithm and the simulation is carried out.

Let us assume that there are 3 types of resources, say, r_1, r_2, and r_3 with 10, 5, and 7 instances of each resource, respectively. These resources r_1, r_2, and r_3 are available at resource centers R1, R2, and R3, respectively. Let the resources be allocated to 5 activities, say, A_0, A_1, A_2, A_3, and A_4. As shown in Figure 7(a), at place A_0, activities have resources (r_1, r_2, r_3)=(0 1 0) initially allocated from resource centers R1, R2, and R3. Similarly, at location A_1, activities have been allocated resources (r_1, r_2, r_3)=(2 0 0) initially. At other places, A_2 has (3 0 2), A_3 has (2 1 1), and A_4 has (0 0 2) resources initially allocated. The additional requirement of resource by each activity is shown in Figure 7(b); i.e., at place A_0, activities have demand of resources (r_1, r_2, r_3)=(7 5 3) from resource centers R1, R2, and R3, respectively. Similarly, at



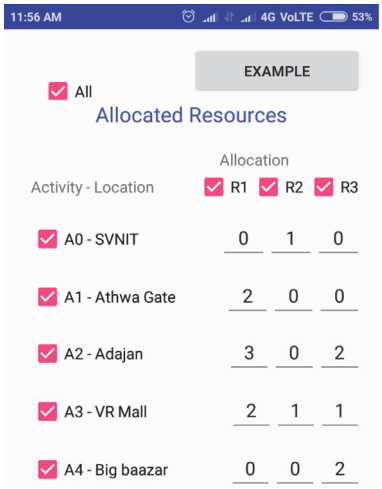
(a) Details of resource centers and emergency locations



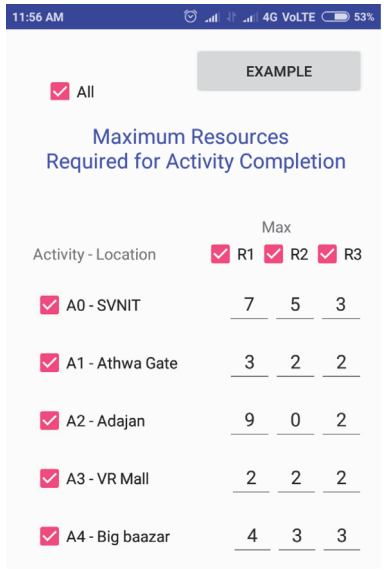
(b) Resource Centers on Google map

(c) Emergency locations on Google map

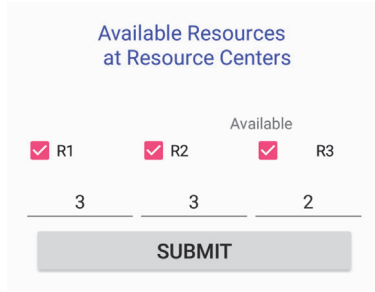
Figure 6: Google Map Representation.



(a) Resource allocation values



(b) Maximum resource required

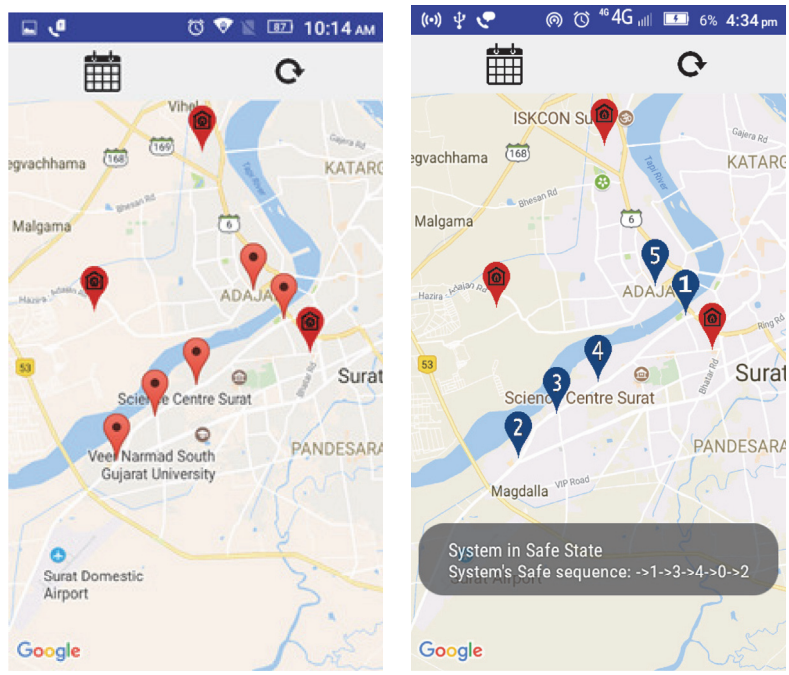


(c) Available resource values

Figure 7: Input details of location and values for all tasks.

other emergency locations, A_1 needs (3 2 2), A_2 needs (9 0 2), A_3 needs (2 2 2), and A_4 needs (4 3 3) additional number of resources from A, B, and C resource centers. After initial allocation, the available numbers of resources at resource centers A, B, and C are 3, 3, and 2, respectively, which is shown in Figure 7(c). The resource centers and emergency locations which are listed in Figure 6(a) are represented using Google map and are displayed in Figure 8(a).

Now, the execution of our proposed resource scheduling algorithm results in resource scheduling of resources for different activities, in a safe sequence, avoiding deadlock and race conditions which are shown in Figure 8(b). Using Banker's algorithm for resource scheduling, the execution of various activities scheduled at different emergency locations is represented by the sequence of numbers 1, 3, 4, 0, and 2 in Figure 8(b). This sequence number represents the emergency location places and without priorities. That is, the activities are scheduled at places A_1, A_3, A_4 A_0 and A_2 in sequence. So activity at place A_1 is executed first followed by the activities at place A_3. Then, activities at place A_4 are executed followed by activities at A_0 and A_2. In terms of actual locations of the Surat city, the resource is allocated to Athwa Gate first,



(a) Resource centers and activity representation

(b) Safe Sequence Scheduling

Figure 8: Scheduling of activities without priorities.

followed by VR Mall, Big-bazaar, SVNIT, and Adajan in sequence. After the allocations of resources at these places in sequence as discussed above the sequence of execution at different places is shown with a blue marker in Figure 8(b).

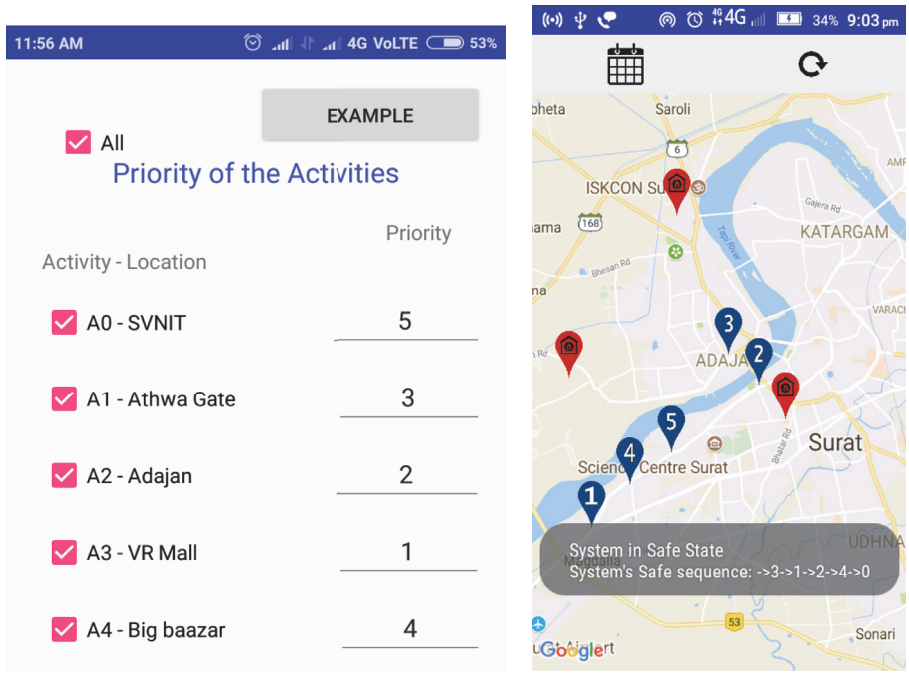
Afterward, the simulation result is presented where another parameter, i.e., the priority of the activity, is incorporated for resource scheduling. The visual representation assignment of the priority to different activities is shown in Figure 9(a). The activities at place A_0 are assigned lowest priority whereas the activities at place A_3 are assigned the highest priority compared to other places. The output of resource scheduling method, with two parameters, the waiting time for availability of the resource for different activities and the priority of the activity, is displayed in Figure 9(b).

The resource scheduling for different activities at different places using the proposed method results in a different sequence of places for executing different activities. The sequence is now, 3, 1, 2, 4, and 0, as shown in Figure 9(b). This sequence indicates that the scheduling of the activities at place A_3 should be executed first due to the higher priority, followed by activities at place A_1. The activities at place A_2 are having higher priority than the activities at place A_3. Because the requested number of resources is 9 from resource center A by activities at place A_2 it is not possible to allocate those many resources as the available number of resources is only 5. Due to nonavailability of the sufficient number of resources, the next priority activities at place A_1 are executed. Afterward, the activities at place A_2, followed by the activities at place A_4 and the activities at place A_0, are executed in sequence. Using this sequence, now, the resources are allocated to VR Mall, followed by Athwa Gate, Adajan, Big-bazaar, and SVNIT places, respectively. The sequence of places for execution of different activities using our resource scheduling method is shown with a blue marker in Figure 9(b). This demonstrates that, using our priority based resource scheduling algorithm, it is possible to execute the different activities at different places without deadlock or any race conditions for availing different resources and it results in a safe sequence of activity execution across different places.

6.3. Comparative Analysis. For demonstrating the efficacy of our proposed algorithm, the performance analysis is carried out by comparing our approaches with different algorithms reported in the literature [10, 35, 36, 50]. The approaches proposed for resource allocation with first come first served (FCFS) strategy [50], greedy strategy [10], stable matching approach [35], and maximum bipartite graph approach [36] are considered for our comparison purpose. The performance is compared using the standard parameters [53], that is, fairness in resource allocation and execution time for completing the different activities.

The proposed approaches are devised in such a way that maximum utilization of the resources is achieved. For each activity, the resources were allocated with maximum fairness. Jain et al. [53] proposed a technique to measure the fairness by quantifying it. The fairness is given by the following equation

$$f(X) = \frac{\left[\sum_{i=1}^{l} a_i\right]^2}{n} \times \sum_{i=1}^{l} a_i^2 \tag{8}$$



(a) Priority values (b) Safe Sequence Scheduling

Figure 9: Scheduling of Activities with Priorities.

where $0 \leq f(X) \leq 1$ is fairness measure of resource allocation. $X = (a_1, a_2, \ldots, a_l)$ represents the number of resources a_i allocated to activity i. There are l activities assumed. A large value of $f(X)$ represents fairer resource allocation from the system perspective. The corresponding result is shown in Figure 10. It is observed that the fairness of Algorithm 1 is better than the other approaches compared. Due to the additional parameter of priority, Algorithm 2 gives less fairness in the allocation of resources than Algorithm 1. From this figure it is clear that the fairness achieved using resource scheduling methods using priority of activities and waiting time for availability of the resource for different activities is much higher compared to that of other methods reported in the literature.

We also analyze the proposed algorithm in terms of computational complexity for critical time analysis and response. As discussed earlier, for the proposed *Resource Scheduling* Algorithm 1, the time complexity is $O(a^2 r)$ with utmost a possible activities and r resources. But brute force approach takes $O(a!/(a - r)!)$. Also, the greedy approach proposed in [10] takes $O(r^2 a \log a)$ because of the sorting of the activities. The time complexity of stable matching approach is $O(ar)$ whereas maximum bipartite graph approach takes $O(pq)$ where $p = |a| + |r|$ and $q = |E|$, i.e., number of edges. The comparative results of computational complexity time analysis are shown in Figure 11, which helps us to determine the proposed algorithm results better than brute force approach [50] which is FCFS and greedy approach [10]. It is observed that stable matching algorithm takes less time followed by simple Banker's algorithm which takes less time among all the approaches compared. Because of the complexity consideration for queuing theory, Algorithm 2 takes more time than Algorithm 1.

7. Conclusion

Resource scheduling and activity management are crucial during disaster circumstances. The functionality of IoT in the IP-network formation for communications in real time monitoring helps in effective tracking and utilization of the available resources. The empirical results show that using queuing theory helps to determine the optimal number of resource centers. Based on the usage of optimum resource centers, the proposed algorithm schedules the resources by considering the priority of the activities depending on the situational requirements. The proposed algorithm is further verified by taking a case study of Surat city and demonstrated using real time android application with the help of Google maps. The proposed approaches are evaluated in terms of fairness and computational complexity time analysis, which shows better results than the approaches reported in the literature. The proposed algorithm can be further extended as future work for dealing with the response time of each resource at respective resource center which, in turn, affects the performance of the overall system.

Appendix

A. Single Server Resource Center

To illustrate the single resource centers with an example, assume the arrival rate of the activity is 8 per hour and service

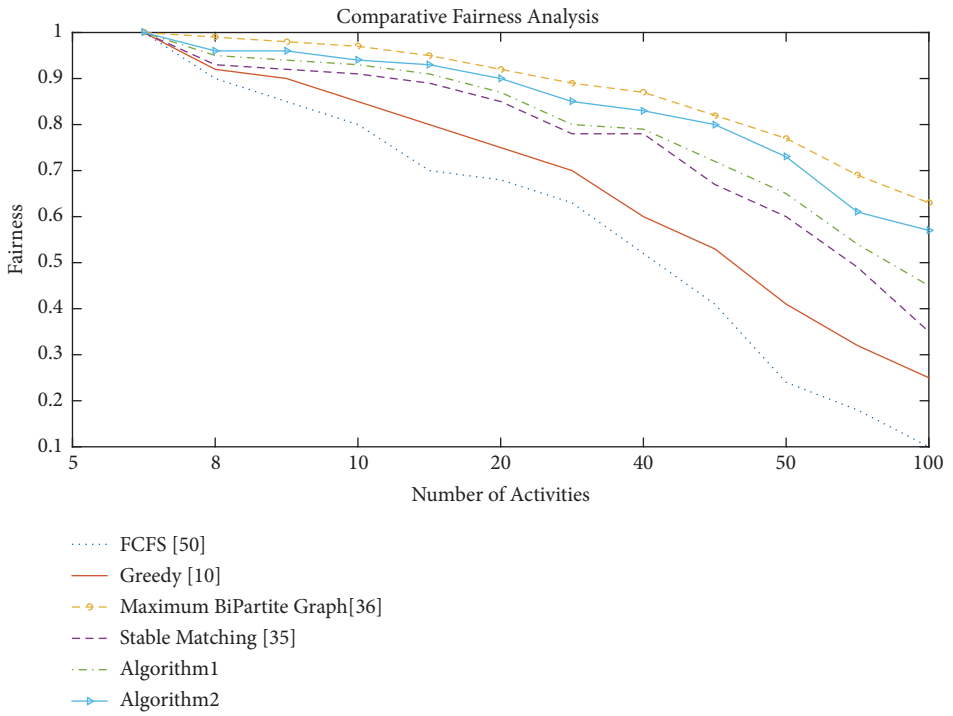


Figure 10: Fairness Comparative Analysis.

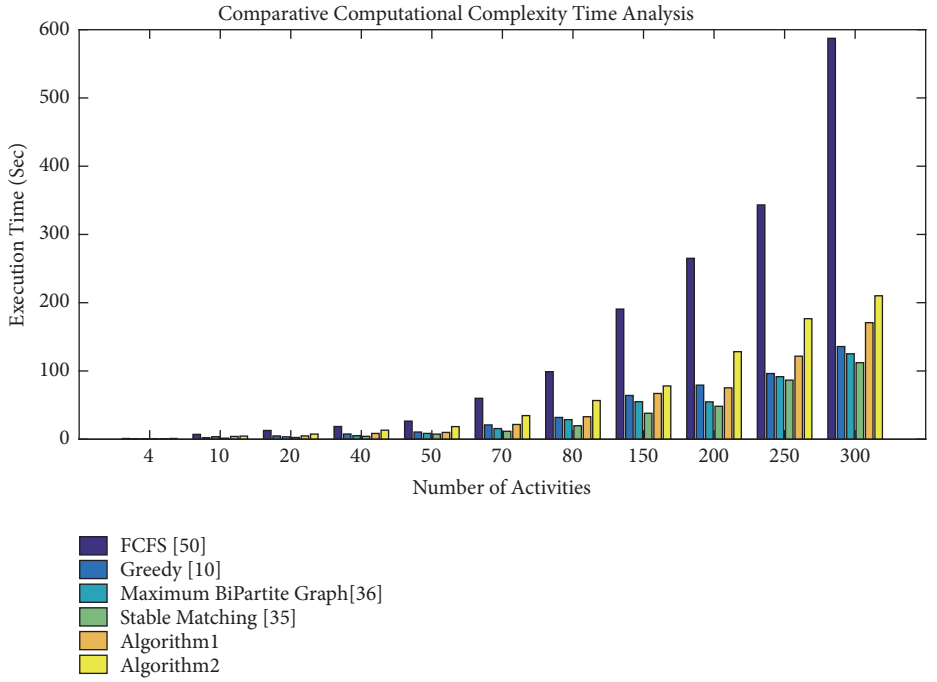


Figure 11: Algorithm Computational Complexity Time Analysis.

rate of resources is given as 9 per hour; then we have the following.

(i) Now, $P_0 = 1 - \rho = 0.111$. 11.1% of the times there is no activity in the system.

(ii) The probability that the resources are not available can be determined by $1 = P_0 = \rho = 0.88$. Therefore, 88.8% of the times the resources are busy.

(iii) The probability that there is no queue is determined by $P_0(no\_queue) = P_0 + P_1 = 0.11 + \rho P_0 = 0.111(1 +$

0.88) = 0.2098. So, 21% of the times there will be no queue.

(iv) The probability of 10 activities in the system, P(10 in the system), $P_{10} = \rho^1 0 P_0 = 0.88^1 0 * 0.11 = 0.0341$. So, 3.4% times, out of 10 activities in the system, 1 is getting served and 9 are waiting for the service.

(v) The probability of at least 2 or more activities in the system is given by $P(n \geq 2) = P_2 + P_3 + \ldots + P_\infty = 1 - P_0 + P_1 = 1 - 0.2098 = 0.7902$. So, 79% of the times there will be 2 or more activities in the system.

(vi) The waiting time of the system is 1 hour.

B. Multiple Server Resource Center

To illustrate the multiple resource centers with an example, assume the arrival rate of the activity is 10 per hour and service rate of resources is given as 6 per hour and the number of resource centers is 2; then, activities arrival rate/(Numner of resource centers * service rate of resources) = $10/12 = 0.833 < 1$.

(i) Now, $P_0 = 0.0909$. So, 9.1% of the times there is no activity in the system.

(ii) The probability that both resources are free such that an activity is assigned is given by $P(Activity\_gets\_a\_resource) = P_0 + P_1 = 0.2424$. So, 24.24% of the times the activity gets the resources.

(iii) The probability that there is no queue is determined by $P_0(no\_queue) = P_0 + P_1 + P_2 = 0.11 + \rho P_0 + \rho^2 P_0/2 = 0.389$. So, 39% of the times there will be no queue.

(iv) Waiting time = 0.6 hr. The waiting time is 36 minutes.

Data Availability

None of the included data is available online.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work is supported by the Ministry of Electronics and Information Technology (MeiTY), funded by Government of India under the Grant No. 13(4)/2016-CC&BT.

References

[1] M. Wahlstrom and D. Guha-Sapir, *The Human Cost of Weather-Related Disasters 1995-2015*, UNISDR, Geneva, Switzerland, 2015, https://www.unisdr.org/files/46796_cop21weatherdisastersreport2015.pdf.

[2] "Data collection Survey for Disaster Prevention in India," Japan, 2015, http://open_jicareport.jica.go.jp/pdf/12245155.pdf.

[3] M. Muaafa, A. L. Concho, and J. E. Ramirez-Marquez, "Emergency Resource Allocation for Disaster Response: An Evolutionary Approach," 2014.

[4] L. Pearce, "Disaster management and community planning, and public participation: how to achieve sustainable hazard mitigation," *Natural Hazards*, vol. 28, pp. 211–228, 2003.

[5] G. M. Lee, N. Crespi, J. K. Choi, and M. Boussard, "Internet of things," in *Evolution of Telecommunication Services*, pp. 257–282, Springer, 2013.

[6] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): a vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[7] L. Yang, S. H. Yang, and L. Plotnick, "How the internet of things technology enhances emergency response operations," *Technological Forecasting & Social Change*, vol. 80, no. 9, pp. 1854–1867, 2013.

[8] A. Meissner, T. Luckenbach, T. Risse, T. Kirste, and H. Kirchner, "Design challenges for an integrated disaster management communication and information system," in *Proceedings of the First IEEE Workshop on Disaster Recovery Networks (DIREN '02)*, vol. 24, pp. 1–7, 2002.

[9] G. Studie, "Market analyse katastrophen-und not fall management systeme," Kaiserslautern, 2002.

[10] L. Özdamar and W. Yi, "Greedy neighborhood search for disaster relief and evacuation logistics," *IEEE Intelligent Systems*, vol. 23, no. 1, pp. 14–23, 2008.

[11] T. J. Cova and R. L. Church, "Modelling community evacuation vulnerability using GIS," *International Journal of Geographical Information Science*, vol. 11, no. 8, pp. 763–784, 1997.

[12] A. M. Caunhye, X. Nie, and S. Pokharel, "Optimization models in emergency logistics: a literature review," *Socio-Economic Planning Sciences*, vol. 46, no. 1, pp. 4–13, 2012.

[13] H. O. Mete and Z. B. Zabinsky, "Stochastic optimization of medical supply location and distribution in disaster management," *International Journal of Production Economics*, vol. 126, no. 1, pp. 76–84, 2010.

[14] G. Barbarosolu, L. Özdamar, and A. Çevik, "An interactive approach for hierarchical analysis of helicopter logistics in disaster relief operations," *European Journal of Operational Research*, vol. 140, no. 1, pp. 118–133, 2002.

[15] E. T. Erdemir, R. Batta, P. A. Rogerson, A. Blatt, and M. Flanigan, "Joint ground and air emergency medical services coverage models: a greedy heuristic solution approach," *European Journal of Operational Research*, vol. 207, no. 2, pp. 736–749, 2010.

[16] E. Rolland, R. A. Patterson, K. Ward, and B. Dodin, "Decision support for disaster management," *Operations Management Research*, vol. 3, no. 1-2, pp. 68–79, 2010.

[17] A. C. Chapman, R. A. Micillo, R. Kota, and N. R. Jennings, "Decentralised dynamic task allocation: a practical game: theoretic approach," in *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, vol. 2, pp. 915–922, 2009.

[18] J. Zhang, M. Zhang, F. Ren, and J. Liu, "An innovation approach for optimal resource allocation in emergency management," *IEEE Transactions on Computers*, 2016.

[19] S. Babashetty, A. G. Kurian, and R. Vonna, "Multi-tenant disaster recovery management system and method for intelligently and optimally allocating computing resources between multiple subscribers," (Jul. 26 2016), uS Patent 9,400,718.

[20] C. Zhang, X. Liu, Y. Jiang, B. Fan, and X. Song, "A two-stage resource allocation model for lifeline systems quick response with vulnerability analysis," *European Journal of Operational Research*, vol. 250, no. 3, pp. 855–864, 2016.

[21] S. Chang, M. Ichikawa, H. Deguchi, and Y. Kanatani, "A general framework of resource allocation optimization and dynamic scheduling," *SICE Journal of Control, Measurement, and System Integration*, vol. 10, no. 2, pp. 77–84, 2017.

[22] J. H. Lambert and C. E. Patterson, "Prioritization of schedule dependencies in hurricane recovery of transportation agency," *Journal of Infrastructure Systems*, vol. 8, no. 3, pp. 103–111, 2002.

[23] F. Fiedrich, F. Gehbauer, and U. Rickers, "Optimized resource allocation for emergency response after earthquake disasters," *Safety Science*, vol. 35, no. 1, pp. 41–57, 2000.

[24] B. Balcik, B. M. Beamon, and K. Smilowitz, "Last mile distribution in humanitarian relief," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 12, no. 2, pp. 51–63, 2008.

[25] F. Wex, G. Schryen, S. Feuerriegel, and D. Neumann, "Emergency response in natural disaster management: allocation and scheduling of rescue units," *European Journal of Operational Research*, vol. 235, no. 3, pp. 697–708, 2014.

[26] M.-S. Chang, Y.-L. Tseng, and J.-W. Chen, "A scenario planning approach for the flood emergency logistics preparation problem under uncertainty," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 6, pp. 737–754, 2007.

[27] J. B. Sheu, "An emergency logistics distribution approach for quick response to urgent relief demand in disasters," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 6, pp. 687–709, 2007.

[28] N. Altay and W. G. I. Green III, "OR/MS research in disaster operations management," *European Journal of Operational Research*, vol. 175, no. 1, pp. 475–493, 2006.

[29] A. M. De Bruin, A. C. Van Rossum, M. C. Visser, and G. M. Koole, "Modeling the emergency cardiac in-patient flow: an application of queuing theory," *Health Care Management Science*, vol. 10, no. 2, pp. 125–137, 2007.

[30] L. Mayhew and D. Smith, "A multi-class queuing network analysis methodology for improving hospital emergency department performance," *Health Care Management Science*, vol. 11, no. 1, pp. 11–21, 2008.

[31] J. K. Cochran and K. T. Roche, "Using queuing theory to analyse the governments 4-h completion time target in accident and emergency departments," *Computers & Operations Research*, vol. 36, no. 5, pp. 1497–1512, 2009.

[32] M. L. McManus, M. C. Long, A. Cooper, and E. Litvak, "Queuing theory accurately models the need for critical care resources," *Anesthesiology*, vol. 100, no. 5, pp. 1271–1276, 2004.

[33] K. Fitzgerald, L. Pelletier, and M. A. Reznek, "A queue-based monte carlo analysis to support decision making for implementation of an emergency department fast track," *Journal of Healthcare Engineering*, vol. 2017, Article ID 6536523, 8 pages, 2017.

[34] G. R. Rodríguez Jáuregui, A. K. González Pérez, S. Hernández González, and M. D. Hernández Ripalda, "Analysis of the emergency service applying the queueing theory," *Contaduría y Administración*, vol. 62, no. 3, pp. 733–745, 2017.

[35] J. S. Kumar, M. A. Zaveri, and M. Choksi, "Activity based resource allocation in IoT for disaster management," in *Proceedings of the International Conference on Future Internet Technologies and Trends*, pp. 215–224, Springer, 2017.

[36] J. S. Kumar and M. A. Zaveri, "Graph-based resource allocation for disaster management in IoT environment," in *Proceedings of the Second International Conference on Advanced Wireless Information, Data, and Communication Technologies*, pp. 1–12, ACM, 2017.

[37] O. Rodrguez-Espndola, P. Albores, and C. Brewster, "Disaster preparedness in humanitarian logistics: a collaborative approach for resource management in floods," *European Journal of Operational Research*, vol. 264, no. 3, pp. 978–993, 2018.

[38] S. Wang, F. Liu, L. Lian, Y. Hong, and H. Chen, "Integrated post-disaster medical assistance team scheduling and relief supply distribution," *International Journal of Logistics Management*, 2018.

[39] H. Gamitl, J. S. Kumar, and M. A. Zaveri, "Resource monitoring and scheduling for post disaster management using internet of things," in *Proceedings of the 3rd IEEE International Conference for Convergence in Technology*, pp. 1–5, 2018.

[40] X. V. Doan and D. Shaw, "Resource allocation when planning for simultaneous disasters," *European Journal of Operational Research*, vol. 274, no. 2, pp. 687–709, 2019.

[41] M. Lawley, S. Reveliotis, and P. Ferreira, "The application and evaluation of banker's algorithm for deadlock-free buffer space allocation in flexible manufacturing systems," *International Journal of Flexible Manufacturing Systems*, vol. 10, no. 1, pp. 73–100, 1998.

[42] H. Madduri and R. Finkel, "Extension of the banker's algorithm for resource allocation in a distributed operating system," *Information Processing Letters*, vol. 19, no. 1, pp. 1–8, 1984.

[43] J. S. Kumar, M. A. Zaveri, and M. Choksi, "Task based resource scheduling in IoT environment for disaster management," *Procedia Computer Science*, vol. 115, pp. 846–852, 2017.

[44] M. A. Zaveri, J. S. Kumar, S. Kumar, and M. Choksi, "Collaborative data processing and resource optimization for post disaster management and surveillance using internet of things," Tech. Rep., Ministry of Electronics and Information Technology, 2016.

[45] S. K. Pandey and M. A. Zaveri, "Localization for collaborative processing in the internet of things framework," in *Proceedings of the Second International Conference on IoT in Urban Space*, pp. 108–110, ACM, 2016.

[46] J. S. Kumar and M. A. Zaveri, "Clustering for collaborative processing in IoT network," in *Proceedings of the Second International Conference on IoT in Urban Space*, pp. 95–97, ACM, 2016.

[47] J. S. Kumar and M. A. Zaveri, "Hierarchical clustering for dynamic and heterogeneous Internet of Things," *Procedia Computer Science*, vol. 93, pp. 276–282, 2016.

[48] J. S. Kumar and M. A. Zaveri, "Graph based clustering for two-tier architecture in Internet of Things," in *Proceedings of the of IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pp. 229–233, 2016.

[49] J. S. Kumar and M. A. Zaveri, "Clustering approaches for pragmatic two-layer IoT architecture," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8739203, 16 pages, 2018.

[50] L. K. Comfort, K. Ko, and A. Zagorecki, "Coordination in rapidly evolving disaster response systems: The role of information," *American Behavioral Scientist*, vol. 48, no. 3, pp. 295–313, 2004.

[51] Android 4.4 APIs, 2018, https://developer.android.com/about/versions/android-4.4.

[52] G. Svennerberg, *Beginning Google Maps API 3*, Apress, 2010.

[53] D. C. R. Jain and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared systems," Technical Report DEC-TR-301, Digital Equipment Corporation, 1984.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 6185806, 16 pages
https://doi.org/10.1155/2019/6185806

Research Article

Multiobjective Based Resource Allocation and Scheduling for Postdisaster Management Using IoT

Meghavi Choksi and Mukesh A. Zaveri

Computer Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat 395007, Gujarat, India

Correspondence should be addressed to Meghavi Choksi; meghavichoksi@gmail.com

Received 27 November 2018; Revised 4 February 2019; Accepted 24 February 2019; Published 17 March 2019

Guest Editor: Maurizio Casoni

Disaster is an uncertain phenomenon that arises due to natural as well as man-made calamities. Disaster often causes a high degree of destruction, especially in a very densely populated region. To handle such a situation, efficient resource management strategies are required. Resource management is the most crucial phase of disaster management. Efficient and in-time allocation of resources is very important; otherwise, it may result in more fatalities. In this context, we propose the resource management algorithm, which deals with both over- and underdemand for resources. Resource management requires efficient resource allocation, and in case of overdemand for resources, it must be followed by resource scheduling. In this paper, we introduce a resource allocation technique which is based on multiple objectives having a different set of constraints. We also propose the resource scheduling algorithm based on various parameters. The proposed algorithm uses multiobjective theory for resource allocation which is followed by the implementation of priority-based scheduling technique, in the case of overdemand for resources. Our proposed methods are compared to the existing approaches in the literature. From the simulation results, it is clear that our methods perform optimum resource allocation and scheduling operations.

1. Introduction

The disasters may be natural calamities or man-made hazards, which results in a huge loss of human lives and infrastructures [1, 2]. Disasters are unpredictable and suddenly occurring events which cause mass destruction. Disasters can be classified into natural disasters, such as earthquake, cyclone, flood, and tsunami, and man-made disasters, such as a nuclear explosion, oil spills, and gas leakage. If emergency situations are raised due to disasters, the rescue and recovery operations must be carried out efficiently and quickly. The rescue and recovery operations need different kinds of resources for performing various tasks. Thus, there is a need for optimal allocation of resources at multiple emergency locations. This is a challenging issue when there are multiple locations of disaster, and at each location, the set of rescue related tasks are needed to be carried out. During the last two decades, a lot of preventive and protective actions have been taken by the respective government to identify the disaster prone areas and prepare immediate action plans for postdisaster situations. However, the disaster is still considered as an unpredictable catastrophe. In this regard, it is very crucial to take quick actions to reduce the number of fatalities and minimize the destruction of infrastructure. There are two phases of disaster management [3–5]: predisaster phase and postdisaster phase, as shown in Figure 1. The predisaster phase deals with mitigation and preparedness. On the other hand, the postdisaster phase deals with the in-time response and quick recovery. Resource management is a part of postdisaster management. The effective prevention and mitigation strategies minimize the risk of hazard and its consequences by taking proactive steps.

Mitigation [7] requires long-term planning and learning from the experiences of earlier disasters. It includes public education, hazard assessment, infrastructure improvement, risk assessment, protection of critical infrastructures, and many more, while preparedness deals with volunteer management, strategy deployment, assessment management, and emergency detection, on a broader level. Disaster response activities include emergency shelter, search and rescue

Figure 1: Phases of a disaster management.

operations, treatment of injured people, damage assessment recovery, resource allocation, situation management, and stabilization. To perform the above-mentioned operations, the rescue team must have the knowledge of demands at disaster places, for which there is a need for coordination and cooperation among various tasks, such as search and rescue, setting up relief camp, and providing initial treatment to injured people. Search and rescue operations may include looking for missing people, providing first-aid, taking severely injured persons to the hospital, and providing them shelter to safe places. These operations are the most crucial and hold the highest priority. In this phase, an emergency warning is given to the common people using various means like media, social networks, and social platforms. Inventory management and vehicle tracking ensure efficient and faster ways for resource allocation. The recovery operation can be elaborated by reconstruction and restoration of services like electricity, communication, transportation, rehabilitation of ecosystem, and redevelopment, as and when required.

In the postdisaster phase, the most crucial task is to collect accurate information from disaster-affected regions. This may be performed by using an aerial surveillance system to collect the relevant data. The use of an aerial system is to obtain information about the ground situation like road network of the disastrous area. In such a situation, the unmanned aerial vehicle is considered to be a very useful and flexible tool for acquiring information about the ground reality. These data may be collected through appropriate sensors and devices, which may be deployed at the disaster sites prior to the disaster occurrence, as a part of mitigation through the aerial system. In the current scenario, the advent of Internet of Things (IoT) has made it possible to collect and communicate the acquired data using Internet and performing various tasks. One of the most important tasks in such situations is resource management. Resource management includes two important tasks, namely, resource allocation and resource scheduling. For instance, in case of an earthquake, several zones in a city may be affected. Over there, resources such as a crane, hydraulic jack, hydraulic cutter, an ambulance, first-aid support, food packets, and clothes are needed. Based on the kind of activities or tasks, various resources are distributed. An optimum resource allocation strategy decides how resources may be distributed from different warehouses to the affected region in time. However, in case of a limited number of resources, the resources must be scheduled to reduce the demand and supply gap during the recovery and rescue operations in the disaster-affected regions.

Resource management requires different objectives to be fulfilled depending upon the set of tasks to be performed. These objectives include various parameters like the cost of operation, technology to support real-time response strategies, satisfying demand in minimum time, and maximizing the utilization of available resources. In addition, there is a need to serve maximum possible affected areas with the available number of resources. This becomes even more challenging when all these objectives are needed to be fulfilled at all the places and at the same time. In such situations, resource scheduling and allocation play a vital role. The resource allocation strategy should allocate the appropriate number of resources at different places as per the demand in such a way that other places may be served with the remaining number of resources optimally. Also, resource scheduling supports the process of resource allocation by aiding efficient scheduling of the resources for the optimal allocation.

The allocation and scheduling operations need effective data collection and communication. If resources are tagged with appropriate sensor devices and RFID, then it becomes easy to avail the real-time information about various resources through an established network and it allows one to perform optimal allocation and scheduling of the resources. Thus, it is very important to set up such a network among sensor devices and it is expected that the network should be wireless and IP-enabled. Due to this, the data or information about the resources can be available from anywhere. This is possible with the help of an Internet of Things (IoT). The IoT helps in achieving the same purpose and it is possible to maintain the real-time record of the resources and disseminate this information where and when required. Internet of Things [8] is a modern emerging technology. It allows communication among things or objects from anywhere and at any time. Wireless Sensor Network (WSN) has limitations like energy constraint, limited communication range, low transmission, and reception power. IoT is an IP-enabled solution.

As part of predisaster activities, the mitigation and preparedness are very important, whereas for postdisaster activities, distributing resources in an efficient way in real time is very important. The distribution of resources must be done carefully and it is required to trace out whether the resources are allotted and utilized efficiently or not. In this process, mitigation or preplanning activities take care of the tagging of resources in advance. Even the suppliers nowadays are using bar code based resource identification with a detail of batch, date, and packaging information. So it is assumed that the resources are tagged with all details in the form of a simple barcode or with RFID tag [9]. While shifting these resources from the warehouse there is no need for manual efforts to maintain the stock of it. The tag reader may help

in this task. The concept of tagging is not only for nonliving objects or materials available for distribution but it may be for human resources having RFID enabled identity card with their skills and other details embedded on it. This information may be traced for utilizing the human resources efficiently. Even without RFID based tagging, some mechanism is required to trace the distribution of different resources. In this context, RFID based tagging helps in monitoring and tracking the resources. Also, in a logistic vehicle, a GPS module is attached on an IoT device. IoT device can be a low-cost standard device available with embedded boards like Raspberry Pi, Arduino, or Intel Galileo. In our proposed system, for demonstration and evaluation purpose, the IoT devices, Raspberry Pi, Arduino, and various sensor motes are used. A few of these motes are embedded with GPS modules as well. GPS sends real-time latitude and longitude to make it possible to track the resources tagged with such devices. Further, the system is also integrated with Google Maps, which shows real-time traffic scenario and road surfaces, which helps in real-time decision making.

This paper proposes a solution for real-time resource allocation and scheduling to handle the multiple objectives in case of postdisaster situations. The proposal first allocates the resources and schedule them to fulfill the demand for completing various tasks or activities at different locations simultaneously. The proposal uses four different aspects: priority of the task, transportation cost, resource utilization index, and scheduling time index, which are discussed later in detail in the proposed algorithm section. All these factors are combined in the proposal using a multiobjective based model for resource management. The optimization is performed using Lingo programming tool [10] and the real-time data and information are mapped using Android-based visual system. The outline of the paper is as follows: related work is described in Section 2. Multiobjective based resource management scheme for allocating and scheduling is proposed in Section 3. In Section 4, results and analysis are described and discussed. This is followed by a conclusion.

2. Related Work

In this section, a review related to resource management is presented. For postdisaster management, it is very crucial to handle resource allocation and its scheduling efficiently and in an optimized way. Managing resources is very difficult in case of various critical situations such as calamities and wars because one has to consider various parameters for decision making so as to reduce the fatalities. For example, in the case of postdisaster management for carrying out rescue and recovery operations, various parameters like distance from resource center, availability of resources, and number of people affected must be considered while allocating resources.

Conventional resource management systems used for postdisaster management are based on Geographical Information System (GIS), which uses geospatial data [11, 12] collected using various satellites. The GIS system compares the postdisaster geospatial data with that of a predisaster situation for assessment of damage occurred in the affected area. The GIS-based system works with the help of satellites. It is mandatory to hire satellite service for the GIS-based system which is expensive. Disaster management approaches using drones have been proposed in [13, 14]. The drone-based system helps in acquiring images of disaster-affected areas, but it does not provide location information along with it, which is generally available using the GIS-based system. To overcome this problem, nowadays drones are available with in-built GPS sensors so that the location information can be made available. Decision support system framework [15] has been proposed for decision making during disaster based on information collected from multiple agencies, previous disaster statistics, and records which were stored in a centralized system. The centralized decision support system [15] has been used for resource dispatching and tracking based on multiple criteria for disaster management. Here, multiple objectives are considered to allocate resources during an emergency situation to all the places that are severely affected by a disaster. Decentralized [16] resource management scheme for postdisaster allows periodic sharing of resource information by each relief camp in the delay tolerant network.

Resource management requires optimal resource allocation during the critical time and it is quite challenging. Resource allocation is the process of distributing the available resources in such a way that the maximum utilization of resources can be achieved. Resource allocation should ensure that each objective must be fulfilled in an optimized way by incorporating multiple constraints and criteria. A significant amount of research has been carried out in the fields of resource allocation for applications like healthcare management [17–20], logistic industries, manufacturing companies, cloud computing [21–23], and supply chain management [24]. Proper utilization of resources based on demand and availability is very important, in the absence of which, underutilization of resources has an adverse effect in terms of achieving the goals for an organization. These resources may include, but not limited to, the medical kits, ambulance, fire vehicles, food packages, and rescue boats. The delivery of these resources in real time definitely reduces the losses. Mohammed Muaafa et al. [25] have used multiobjective resource allocation scheme for a medical emergency response. This scheme considers a number of objectives like the location of an emergency unit and the number of victims at each unit. Multiobjective based resource allocation has been used in operational research fields, which mainly deals with decision making based on multiple criteria at a given time. F. Fedrick et al. [26] have described an optimized resource allocation in an emergency situation for handling earthquake-like situation. It focuses on optimized resource allocation for the search and rescue period.

In case of a limited number of resources, along with resource allocation, there is a need for resource scheduling for providing required resources to different tasks or activities which must be completed over a predefined time. The scarcity of resources may delay the relief work and may create adverse situations while performing postdisaster activities. Resource scheduling is needed when the number of available resources is less and the demand is more. Resource should be scheduled in such a way that all tasks must be completed in minimum time and there is a minimum loss of human lives.

Conventional resource scheduling approaches are generally divided into preemptive [27] or nonpreemptive [28, 29] approaches, respectively.

During postdisaster management, different tasks need to be scheduled. Various tasks may have different priorities. In priority-based task scheduling, each task is assigned a priority. Thus, the more important task is executed first. Naturally, priority-based task scheduling [30] provides better performance while dealing with postdisaster activities [31]. While allocating and scheduling the different resources, it may happen that the deadlock situation arises and starvation for resource occurs. The Banker's algorithm has been used for avoiding the deadlock for different applications [32–34]. In [32], the Banker's algorithm has been used for ensuring finite turnaround time for different jobs. The disadvantage of this approach is that it does not take care of resource utilization. In case of critical response and recovery operations, it is mandatory to have an efficient resource utilization due to a limited number of resources and a large number of regions needing the relief operation. In this view, the method proposed in this paper considers both issues, the scheduling and utilization of resources, simultaneously. The approach based on Banker's algorithm has been described in [33] for improving utilization of automatic guided vehicles through the proposed scheduling policy. In a real-time application, the demand and availability of resources should be considered for scheduling and allocating the resources. Our proposed method considers the demand and availability of resources while scheduling and allocating them to a different task in an optimized way. In [34], resource provisioning algorithm based on Banker's algorithm has been described for multi-access edge computing. In this paper also the main goal is to avoid a deadlock situation. It also considers overdemand and delays while provisioning the resources. In disaster management, there are several tasks to be completed based on an emergency, the tasks must be performed in particular order, and accordingly resources must be scheduled and allocated. Our proposed algorithm takes care of this by assigning the priority to each task, and priority-based resource allocation and scheduling help in handling several tasks efficiently.

As discussed in the literature survey, a number of methods have been reported for scheduling jobs and utilizing the resources. In the case of a limited number of resources, the resource scheduling is very much necessary. In this context, there is a need for developing the method which takes care of various important tasks and performs the allocation and scheduling of resources by considering various objectives and criteria. There are a number of issues like executing the tasks based on their importance, completing all task as early as possible, and completing all tasks with minimum cost. In this paper, we take care of these issues, which differentiates our proposed method from other techniques reported in the literature. We propose multiple objectives based resource allocating and scheduling algorithm for performing different tasks on hand with priority. The proposed method is evaluated based on the completion time of different tasks and the utilization of resources. Overall, using our proposed algorithm, the performance of the system is improved in terms of the number of parameters, namely, fair allocation of resources, utilization of resources, and time for completing a particular task at the different places simultaneously.

3. Proposed Resource Management Approach

In this section, the proposed approach for resource management is presented. The novelty of the proposed algorithm is as follow. From discussion in the literature survey, there is a need for an efficient method for resource allocation and utilization. Several tasks need to be completed based on their importance in a particular order. So assigning the priority to each task and accordingly allocating and scheduling the resources are very important. For achieving this, there may be several objectives that must be fulfilled. In this context, this paper introduces the novel algorithm for resource allocation and scheduling based on priority using multiobjective based criteria for completing the different tasks efficiently. When a disaster occurs, multiple places are affected. To handle such a situation, the rescue and recovery operations must be carried out in an effective manner. For rescue and recovery operations, various activities or tasks must be planned and executed accordingly. For executing these activities or tasks, various types of resources are required for completion. So there is a need to distribute the number of resources to these tasks or activities in time. Each task or activity may need the different type of resources and, hence, it is important to allocate the required number of resources at each place for completion of the tasks or activities on time, which may help to minimize fatality. The above discussion may be visualized through a postdisaster management workflow as depicted in Figure 2. Assume there are N disaster-affected places. These places are represented as a set $\mathscr{P}$ given by

$$\mathscr{P} = \{P_1, P_2, P_3, \ldots, P_N\} \tag{1}$$

where P_i is i^{th} place in the affected area for $i = 1, \ldots, N$. Assume that there are M number of warehouse centers represented by $\mathscr{W}$ such that

$$\mathscr{W} = \{W_1, W_2, W_3, \ldots, W_M\} \tag{2}$$

where W_j is specific j^{th} warehouse located on the field where relief work is carried out for $j = 1, \ldots, M$. The warehouse center is basically a storage house for different resources. Each warehouse has K different types of resources, where K varies from one warehouse to another. The number of these resources are represented as a set $\{R_1, R_2, R_3, \ldots, R_K\}$ at a particular warehouse. For representing the number of resources of a particular type located at a specific warehouse, the symbol R_{jk} is used to indicate the number of resources of type k located at j^{th} warehouse.

At each disaster place, there is a need to carry out certain number of tasks. These tasks are denoted by $\{T_1, T_2, T_3, \ldots T_n\}$. To carry out these tasks, different types of resources are required. There is a possibility that the demand at different places may be more as compared to the available number of resources, or it may be less than the total number of available resources. In brief, there may be an overdemand or underdemand for resources. The

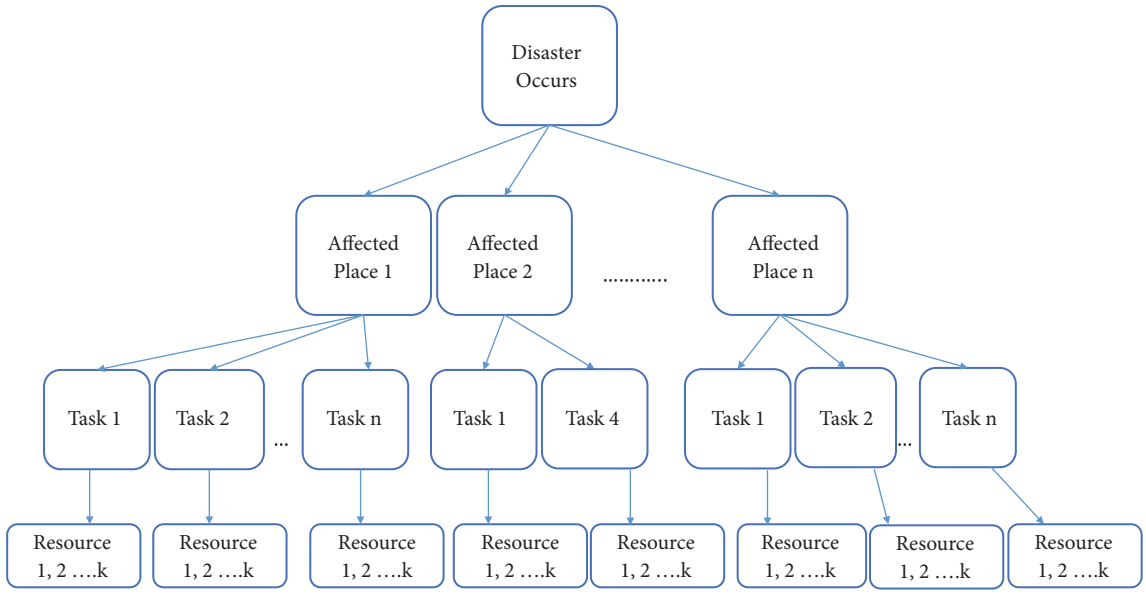


Figure 2: Disaster management workflow.

proposed method for resource management must consider this aspect. In case of overdemand for different resources, i.e., a demand for resource is more than its availability, the resource management process is divided into two phases: *resource allocation* followed by *resource scheduling*, while in case of underdemand for resources there is only one phase, i.e., *resource allocation*. The demand for a particular resource of type k at place i is D_{ik} and the number of resources of type k available at warehouse j is R_{jk}. Using these notations, overdemand and underdemand for resources are depicted in the form of equations shown below. Equations (3) and (4) depict the overdemand and underdemand, respectively.

$$\sum_{i=1}^{N}\sum_{k=1}^{K} D_{ik} \geq \sum_{j=1}^{M}\sum_{k=1}^{K} R_{jk} \tag{3}$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} D_{ik} < \sum_{j=1}^{M}\sum_{k=1}^{K} R_{jk} \tag{4}$$

Based on the conditions mentioned in (3) and (4), both activities, namely, resource allocation and scheduling, need to be carried out, or only resource allocation is performed.

3.1. Resource Allocation. Resource allocation is a systematic approach for assigning the resources as per the requirement. As mentioned earlier, postdisaster management needs to accomplish various tasks for fulfilling different objectives set as part of the rescue and recovery operation that is carried out. In this context, it is desirable to have a method which takes care of these objectives while allocating resources. In this view, we propose a multiobjective based resource allocation scheme. Multiobjective based allocation scheme decides how many resources must be allocated from which warehouse to the disaster place. Naturally, the proposed scheme should allocate resources in the first phase in such a way that the number of resources should be distributed to those places having maximum requirement and priority of an activity or tasks to be performed. In this view, for fulfilling multiple objectives, we introduce various parameters. These parameters help in defining the multiobjective functions for accomplishing various tasks of rescue and recovery operation.

For rescue and recovery operation, it is very important that various tasks or activities must be carried out in time to reduce postdisaster fatalities. In this context, we consider mainly four parameters. There are a number of tasks to be performed and each task may have a different amount of urgency. It is preferable to assign priority to different tasks. So the first parameter is proportional priority. Second, the resources should be availed in minimum time at the required place. The resources may be distributed from various warehouses located at different places as stated earlier. So the second parameter is the transportation cost. The third parameter is resource utilization as each resource must be utilized to its maximum capacity. In the case of a limited number of resources, the scheduling is required so that the resources will be available for different tasks wherever it is required. In this view, the fourth parameter is scheduling time. These parameters are defined and discussed in detail as follows.

3.1.1. Proportional Priority. For handling the postdisaster situation, the main goal is to reduce fatality. For example, there is a mass destruction due to the collapse of several buildings in disaster-affected region, there are a number of people injured at different locations, and now a number of ambulance vehicles are required for shifting these injured people from different places to the nearby medical center for treatment. In such a situation, these places need to be served

first with high priority to reduce fatality. As there may be such several places and the number of injured people may vary from one place to another, it is required that carrying vehicle should be available at a particular place first, where there is more number of injured people compared to the place where there is less number of injured people. This aspect can be incorporated by assigning the priority proportional to the requirement of the resources. For algorithmic description, the number of injured people who need the service of an ambulance is represented as an entity count. This entity count needs a service of a particular resource. That is, *ECS* (entity count for a service or resource) represents entity count as the number of injured people for the example cited here. For assigning proportional priority to a particular service at specific location i, $Priority_i$ is defined as

$$Priority_i = \frac{ECS_i}{\sum_{i=1}^{i=N} ECS_i} \tag{5}$$

where $Priority_i$ is priority value at disaster-affected place P_i and ECS_i represents the entity count at location i or place P_i that needs the service of a particular resource. The proportional priority is normalized with respect to the count of entities at all places in the disaster-affected region.

3.1.2. Transportation Cost. During rescue and recovery operations, the resources are distributed from warehouses to different places in the disaster-affected area. To reduce the fatality, the resources should be provided as quickly as possible. It is quite possible that the roads may be damaged due to the disaster. So identifying the route to the desired location for supplying the resources in shortest span of time is mandatory. The route identified should be such that the transportation time is minimum. In this context, in our proposed algorithm, the transportation cost is incorporated as one of the parameters. The transportation cost is evaluated based on the time and existence of the route to desired location. Drones, developed under the research grant as mentioned in the Acknowledgments, help in capturing images and hence identifying the existence of the route. In case of availability of network access, our proposed system identifies this route through Google Maps.

The existence of the route is represented as reachability between disaster-affected place and warehouse. The reachability is set to 1, if a route exists; otherwise it is set to 0, as shown in (6).

$$Reach_{ij} = \begin{cases} 1, & \text{if route exists between place } P_i \\ & \text{and warehouse } W_j \\ 0. & \text{Otherwise} \end{cases} \tag{6}$$

The objective is to minimize the transportation cost so that the resources should be available in minimum time to the disaster-affected places. Based on this assumption, the objective is defined as

$$\min \sum_{i=1}^{N} \sum_{j=1}^{M} transportation\_cost_{ij} * Reach_{ij} \tag{7}$$

where the $transportation\_cost_{ij}$ is measured in terms of time and distance between the affected place P_i and the warehouse W_j.

3.1.3. Resource Utilization. Another parameter chosen for defining multiobjective based problem formulation is based on the utilization of resources. For rescue and recovery operations, the availability of resources is always critical. It is desired that various resources should be utilized maximally. In this context, the following objective is maximized for achieving maximum resource utilization. The same criteria may be stated in a different view, i.e., serve the maximum number of disaster-affected places with a minimum number of resources. Indirectly, each resource utilization is maximal. Maximal resource utilization can be achieved considering the demand for resources of type k, at affected place i, and with available number of resources that should be allocated at place i for a specific task. It may happen that there are a number of tasks to be performed at place i. Each of these tasks may have a different priority, i.e., proportional priority defined in (5). Thus, some of the tasks may have higher priority compared to other tasks at the same place. The number of such tasks may also vary from one place to another place. In this view, the different places are also assigned priority given by *Place_Priority*, which are defined in terms of the number of tasks with higher priority. If the number of tasks requiring resource type k has high proportional priority, then the *Place_Priority* is higher for that resource type. The *Place_Priority* at place i for resource type k is defined as $Place\_Priority_{ki}$. The resource utilization is measured in terms of available resources based on the number of tasks with their respective priorities which need these resources and the priority of place are defined as above. The objective is to maximize resource utilization using the following equation:

$$\max \frac{\sum_{i=1}^{N} \sum_{k=1}^{K} D_{ik}}{\sum_{j=1}^{M} \sum_{k=1}^{K} R_{jk}} * Priority_i * Place\_Priority_{ki} \tag{8}$$

where $Place\_Priority_{ki}$ is a priority of place i for resource type k.

3.1.4. Scheduling Time. It is important to accomplish various tasks of rescue and recovery operations in time. These tasks need different resources. It may happen that the resources may be limited in number. So, it is required to schedule different tasks and allocate the resources to these tasks. Indirectly various resources should be scheduled for different tasks in a way that all tasks may be completed in time. It is also important that the scheduling time should be minimum. In this view, minimizing overall scheduling time is one of the parameters that our algorithm achieves through multiobjective based formulation. In our proposed system, the program evaluation and review technique (PERT) [35, 36] is used to analyze scheduling time, so that in minimum time more number of places can be served and the required number of tasks may be completed. So the objective is to minimize the overall scheduling time at place i. This

minimum time value, $Time_i$ [35], is given by the following equation:

$$\min Time_i = \frac{[O_t + 4M_t + P_t]}{6} \tag{9}$$

where O_t is an optimistic time to accomplish a particular task assuming that everything goes well on time as normally expected, i.e., the best condition. $4M_t$ is the most likely time, i.e., the best estimated time in which task may be completed, and P_t is the pessimistic time which is the maximum possible time taken to complete task considering the worst scenario, i.e., when everything goes wrong.

Multiobjective Based Allocation. Using the four parameters described above, namely, proportional priority, transportation cost, resource utilization, and resource scheduling time, the multiobjective function is evaluated. The multiobjective function provides an optimal value of the decision variable, represented as $Volume_{ikj}$. The decision variable $Volume_{ikj}$ is the number of resources of type k to be allocated from a particular warehouse j at location i to accomplish a particular task based on its priority in minimum time. Formally, the decision variable is defined as

$$\begin{aligned} &Volume_{ikj} \\ &\quad = \min(\text{transportation cost}, \text{scheduling time}) \\ &\qquad + \max(\text{resource utilization}) \end{aligned} \tag{10}$$

subject to

$$Volume_{ikj} <= D_{ik} \tag{11}$$

$$Volume_{ikj} <= R_{jk}. \tag{12}$$

It is important to note that the number of resources of type k that can be allotted at place i from warehouse j should not be greater than the number of resources of type k available at warehouse j. Similarly, the number of resources of type k allotted for a particular task at place i should not be greater than the demand for the same resource at the same place. Both these conditions are defined as constraints in (10) of multiobjective based evaluation.

3.2. Resource Scheduling. As discussed earlier when the demand for resource is more as compared to the availability of resources, the scheduling of the resources is required for completing various tasks. The resource scheduling needs to be carried out considering various parameters and the availability of resources so that different tasks must be completed in time as per their importance. In disaster management, resources are scheduled based on the priority of tasks and urgency of a place in the required time. In this context, it is important that the places which are not served yet need to be served first as part of resource scheduling to avoid starvation of resources or delay in rescue and recovery operations. For our proposed algorithm, for formulating the problem based on multiobjective based criteria, we use TOPSIS [37–39]. TOPSIS provides the ranking based on a set of parameters for scheduling the resources at nonallocated places for completing various tasks. These parameters are considered as scheduling parameters or criteria. The values of these parameters are of two types, classified as benefit parameters and cost parameters. The number of benefit parameters and cost parameters may vary from one application to another. The benefit parameters are set to the number of entity count (ECS) of a different category. For postdisaster management scenario, there may be a scenario where the number of injured people should be shifted from the disaster location to nearby medical center through ambulance; then $parameter_i$ is the count of entities of a particular age group. The idea is to provide service to a lower age group first compared to the higher age group of people. Similarly, the cost parameters are defined. For postdisaster management, in our case, two cost parameters are defined, namely, the evacuation time and transportation cost. Using these parameters, TOPSIS calculates the ranking for nonallocated places for scheduling the resources. In this context, various variables and symbols used by TOPSIS are described below for narrating the proposed resource scheduling algorithm.

Notations

(1) *allocated_places*: Set of places where resources are already allocated.

(2) *non_allocated_places*: Set of unserved places, where resource demand is not satisfied or partially satisfied.

(3) $Scheduling\_index_i$: Scheduling rank calculated for each unserved places.

(4) $need_{ik}$: Need of resource type k at disaster place i.

(5) $Parameter = \{parameter_1, parameter_2, \ldots, parameter_m\}$ is a set of scheduling parameters or criteria as discussed above.

(6) Using the above parameter values at different disaster location, scheduling index matrix α is created which is called the decision matrix as it is used for scheduling decision.

$$\alpha_{ij} = \text{Non\_allocated\_places} \begin{matrix} \\ \\ \\ \end{matrix} \overset{\text{parameters}}{\begin{bmatrix} \alpha 11 & \ldots & \alpha 1m \\ \vdots & \ddots & \vdots \\ \alpha n1 & \ldots & \alpha nm \end{bmatrix}} \tag{13}$$

α is the decision matrix for scheduling index.

Each entry of α matrix is α_{nm}, indicating the value of parameter m which is used for resource scheduling at nonserved place n.

(7) Each parameter described above is assigned a weight value based on the urgency of the group represented by particular parameter. For this weight vector W is defined.

$W = \{w_1, w_2, \ldots, w_M\}$ is weight vector, where w_m is a weight assigned to parameter $parameter_m$. These

weight values are assigned for the above parameter set at each nonserved place n.

(8) Using the above weights, the β matrix is defined where each entry β_{nm} is normalized weight value, i.e., for parameter m at nonserved place n.

(9) $Available_{jk}$: Currently available resource of type k at warehouse j.

(10) $Allocation_{ik}$: Resource of type k allocated to disaster place i.

(11) I': A set of values associated with benefit parameter. The benefit parameter may be the number of entities served as defined above. In general, benefit parameters are those parameters which are associated with an entity that is benefited in terms of different services.

(12) I'': A set of values associated with cost parameter which is described above. The time of completion of a particular task and the distance from warehouse to an affected place are such parameters which affect the completion of different tasks performed as a part of rescue and recovery operation.

(13) Dt_i^+: Positive Euclidean distance from ideal solution.

(14) Dt_i^-: Negative Euclidean distance from ideal solution.

Equation (14) denotes the set of places which are not served. These are places where the resources are not allocated yet or a partial number of resources are allocated and, hence, these places need to be served as soon as possible. So these are places still in need of resources.

$$non\_allocated\_places = \mathscr{P} - allocated\_places \tag{14}$$

After a phase of resource allocation as per the multiobjective function in (10) as described in the previous section, it may happen that there are a number of places left, where the particular types of resources are not allocated due to the limited number of resources. For representing the need for a particular resource of type k at place i, the following variable is defined by (15).

$$need_{ik} = D_{ik} - Volume_{ikj} \tag{15}$$

The above equation calculates the quantity of resources of type k still needed at disaster places i.

Using the multiobjective based resource allocation method as mentioned above, some places are served but there are unserved places where still resources are needed to be allocated. For handling such situations, using TOPSIS technique as described above, *Scheduling_index* is calculated and assigned to each of nonallocated places, considering multiple parameters.

Steps for Evaluating Scheduling_Index

(i) Step 1: Normalize Decision matrix.

The decision matrix is normalized using standard technique mentioned as below.

$$r_{mn} = \frac{\alpha_{mn}}{\sum_{m=1}^{M}\left(\left(\alpha_{mn}\right)^{1/2}\right)} \tag{16}$$

(ii) Step 2: Construct weighted matrix.

$$\beta_{mn} = w_m * r_{mn} \tag{17}$$

(iii) Step 3: Determine positive and negative ideal solution.

(a) Positive ideal solution identifies the best ideal solution by maximizing benefit parameters and minimizing the cost parameters.

$$\begin{aligned} A^+ &= \{\beta_1^+, \beta_2^+, \ldots, \beta_n^+\} \\ &= \left\{\left(\max \beta_{mn} \mid m \in I'\right), \left(\min \beta_{mn} \mid m \in I''\right)\right\} \end{aligned} \tag{18}$$

(b) Negative ideal solution maximizes the cost parameter and minimizes the benefit parameter.

$$\begin{aligned} A^- &= \{\beta_1^-, \beta_2^-, \ldots, \beta_n^-\} \\ &= \left\{\left(\min \beta_{mn} \mid m \in I'\right), \left(\max \beta_{mn} \mid m \in I''\right)\right\} \end{aligned} \tag{19}$$

(iv) Step 4: Calculate separation measure using Euclidean distance. Here, Euclidean distance from a positive solution to the ideal solution is calculated as well as from negative solution to ideal solution. The ideal solution is the solution which considers maximum satisfaction of all criteria, but such a solution is not feasible in a restricted environment. The separation measures are calculated as below using positive and negative ideal solutions.

$$Dt_i^+ = \sqrt{\sum_{m=1}^{M}\left(\beta_{mn} - \beta_n^+\right)^2} \tag{20}$$

$$Dt_i^- = \sqrt{\sum_{m=1}^{M}\left(\beta_{mn} - \beta_n^-\right)^2} \tag{21}$$

(v) Step 5: Calculate $Scheduling\_index_n$.

The $Scheduling\_index_n$ represents the relative closeness to ideal solution.

$$Scheduling\_index_n = \frac{Dt_n^-}{Dt_n^+} + Dt_n^- \tag{22}$$

Arrange *Scheduling_index* in descending order.

The pseudo code of proposed resource management approach and resource scheduling approach is mentioned in Algorithms 1 and 2, respectively.

The resources must be allocated and scheduled efficiently. The demand for different resources at each place must be satisfied without any deadlock situation and starvation. Hence, scheduling should be done in such a way that different tasks must be completed in order of their importance. Moreover, it is desirable that all resources must be utilized maximally. In this context, the demand satisfaction, percentage utilization of resources, and overall completion of different

(1) **Input**: Set of Disaster Places: P_i, Set of Warehouse Centers: W_j, Quantity of Resource Type k at Warehouse Center j: R_{jk}, Demand of resource type k at each disaster place i: D_{ik}
(2) **Output**: Quantity of resource of type k that should be supplied from warehouse Center to disaster place.
(3) **Procedure**
(4) **if** $\sum_{i=1}^{N} \sum_{k=1}^{K} D_{ik} \geq \sum_{j=1}^{M} \sum_{k=1}^{K} R_{jk}$ **then**
(5) Resource_Allocation()
(6) Resource_Scheduling()
(7) **else**
(8) Resource_Allocation()

Algorithm 1: Resource management (RM).

(1) **Input** Number of warehouse centers, Available resources at warehouse centers, Allocated resources to task, $Scheduling\_index_i$
(2) **Output** Scheduling of resource for a particular task at all non-allocated places
(3) Step 2: Sort the $Scheduling\_index_i$ in descending order.
(4) Step 3: Allocate the resources to all $non\_allocated\_places$ as per order of $Scheduling\_index_i$.
(5) **while** till every $non\_allocated\_places_n$ has served resources **do**
(6) **if** $Need_{ik} \geq Available_{jk}$ **then**
(7) Go for next $non\_allocated\_places$
(8) **else**
(9) ▷ Verifying the availability for the resource type k
(10) $Available_{jk} = Available_{jk} - Need_{ik}$
(11) ▷ Allocating the requested resources for $non\_allocated\_places$ i
(12) $Allocation_{ik} = Allocation_{ik} + Need_{ik}$
(13) Add the served resources
(14) Return resource schedule

Algorithm 2: Resource scheduling.

tasks are considered as performance measure parameters. These parameters help in comparing our proposed method with other existing methods. In the simulation results as mentioned in Section 4, these three parameters are evaluated compared with the existing approaches.

4. Result and Analysis

For evaluating the proposed method, the scenario is synthesized using the number of actual places of our city, Surat, located in Gujarat state in India. The disaster places are demonstrated through Google Maps using the Android application developed for visualizing the outcome using our approach. As pointed out earlier, for postdisaster management, completion of various tasks in particular order in minimum time is very important. For completion of these tasks, resources must be allocated and scheduled accordingly. At the same time due to a limited number of resources, it is desired that available resources must be utilized optimally. In this context, the proposed method is evaluated with respect to three parameters, demand satisfaction, percentage utilization of resource, and overall completion time of different tasks altogether. The demand satisfaction means that a percentage of demand for a particular resource type is satisfied or provided at all the places. Percentage utilization of resource depicts how efficiently the algorithm utilizes the resources. With proposed resource allocation and scheduling, it is desired that the overall completion time of different task must be minimum or fast enough compared to the normal sequential way of completing the different tasks.

The proposed multiobjective based resource allocation and scheduling algorithms are implemented in LINGO [10]. The results are visualized using an Android application using Google Maps. For demonstration, our local city, Surat, region is considered. The complete Surat city region is divided into seven zones by Surat Municipal Corporation (SMC). These regions are shown in Figure 3. For simulating and visualization of the proposed algorithm, it is assumed that various parts of different zones are affected by some massive natural disaster. It is assumed that there are seven warehouse centers, one in each zone. There are ten different areas or places affected by the disaster. At each place, the number of tasks to be performed as part of rescue and recovery operation is decided by the management authority. There is a demand for a certain type of resource at each place. For instance, these resources may be an ambulance, fire vehicle, emergency vehicle, rescue boat, food packages, medical container, etc. For example, there is a demand for a particular type of resource, say, ambulance. There are number of injured people, which are entities, requiring the service of shifting to nearby place by the medical center from disaster-affected place. For such a typical scenario, the demand for particular resource at different places with their priority along with entity count is shown in the Table 1.

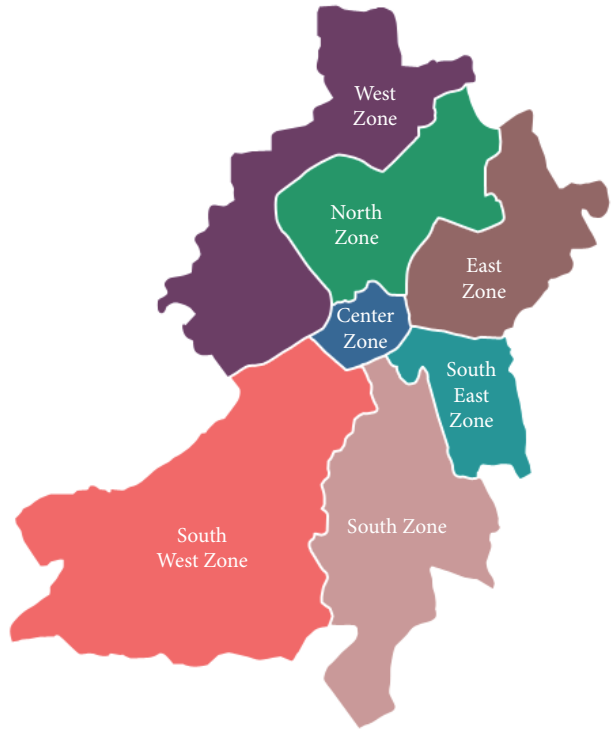


Figure 3: Zone-wise division of Surat city [6].

Table 1: Demand of resource, ECS, and *Place_Priority*.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
Demand at each place D_{ik}	100	37	22	32	41	32	43	38	50	70
Entity Count requires Service (ECS)	400	50	40	300	406	500	470	600	600	800
Place_Priority	1	1	1	2	1	1	2	2	2	2

Table 2: Quantity of resources at each warehouse.

Warehouse Center	W_1	W_2	W_3	W_4	W_5	W_6	W_7
Quantity of Resources R_{jk}	60	55	51	43	41	52	20

Table 3: Route (reachability) from disaster affected place to warehouse center.

$Reach_{ij}$	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
W_1	0	1	1	1	1	1	1	1	1	0
W_2	1	1	1	1	1	1	0	1	1	1
W_3	1	1	1	1	1	1	1	1	1	1
W_4	1	1	0	1	1	0	1	1	1	0
W_5	1	1	1	1	1	0	1	1	1	1
W_6	0	0	0	1	0	1	1	1	1	1
W_7	0	0	0	1	0	1	1	1	1	1

It is important to note that the scenario described here is a very typical one considering all types of situations. In Table 1, at place P_1 the demand is more but ECS is less compared to other places like P_4 to P_{10}. In the given scenario, all entities are injured people who need medical service. Further, all these entities are distributed over different age groups. Our allocation and scheduling algorithm exploits this information while allocating the resources at different places. As mentioned in the algorithm description, the allocation of different resources is based on an evaluation of multiobjective function using different parameters. One of the parameters is a place priority which is based on ECS. The demand is decided by the urgency of resource requirement. As entities are distributed over different age groups, the urgency may vary from one group to another. So it may happen that ambulance service is not required by all groups of entities but only a few entities among these groups require an ambulance service, which are severely injured, and hence the demand is less though ECS is more at places like P_4 to P_{10}.

As mentioned earlier, each warehouse center has different types of resources. The number of resources of type k at warehouse center j is represented by R_{jk}. In Table 2, the number of resources of a particular type available at various warehouse center is mentioned.

The reachability from each disaster-affected places to the warehouse center is shown in Table 3. If a route exists from the warehouse center to disaster-affected place, then it is

indicated by 1. If a route is damaged or having congestion, then it is not possible to reach the place through this route and, hence, it is indicated by 0.

In Figure 4, the distance from disaster-affected place to nearby warehouse center is shown using Google Maps for visualization. Here, latitude and longitude are mapped in a 2-dimensional coordinate system to calculate the distance for further processing. In Figure 4, the upper part of image represents latitude and longitude on y-axis and x-axis, respectively. The place and warehouse are indicated by red dots whereas the distance among these place and warehouse in meters is displayed in the lower part of the same image with blue dots for place and warehouse.

First, the result of resource allocation is described and it is followed by the result and analysis of the resource scheduling. Multiobjective approach is implemented in LINGO [10] and compared to the existing approach [15] for resource allocation. Figure 5 displays the comparison of allocation of a particular type of resource at different places using our proposed approach in this paper and existing approach [15]. In Figure 5, the demand for a particular resource at each disaster-affected place is shown using yellow color. The resource allocated by our proposed approach, which is the value of decision variable $Volume_{ikj}$ derived using (10), is shown with magenta color. The resource allocation by the existing approach [15] is displayed using green color. In [15], for allocating the resources, it considers priority and cost matrix, also considering the number of injured people. In our case, multiple objectives are considered as mentioned in (10) and have used proportional priority for the places, based on entity count (ECS) which requires service. We have also considered the scenario of overdemand as well as underdemand for resources. Further, it is worth mentioning that the category of ECS based on the age group to determine the urgency is also incorporated to evaluate place priority along with other objectives as described earlier.

The scenario considered here for evaluating our algorithm has 10 places P_1 to P_{10} as discussed earlier in this section. The two-level priorities assigned are high and low. High priority is marked with 2 and low with 1 in Table 1. It also depicts the demand, i.e., the number of resources of a particular type, at places P_1 to P_{10}. The places P_4, P_7, P_8, P_9, and P_{10} are having higher priority compared to other places. From Figure 5 it is clear that the proposed approach allocates more resources to those places where priority is high. As mentioned in Table 1, the places such as $\{P_2; P_3; P_5\}$ have less priority and ECS count is also low. As these places are not that crucial, thus during phase 1, i.e., resource allocation phase, no resources are allocated to these places. At places $\{P_1; P_4; P_6; P_7\}$ partial resources are allocated, while at other places $\{P_8; P_9; P_{10}\}$ complete demand is satisfied, as these places are more important and crucial. That is, these places are served first. From simulation results, it is clear that our proposed approach serves the most affected places, where the entity count and the place priority are higher along with other objective parameters described earlier.

The performance of the resource allocation is evaluated using the demand satisfaction which indicates the total number of tasks completed, and resources are served as per demand across the different places. The demand satisfaction is calculated using (23).

$$demand\_satisfaction = \frac{\sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{j=1}^{M} Volume_{ikj}}{\sum_{i=1}^{N} \sum_{k=1}^{K} D_{ik}} \quad (23)$$

Using our proposed approach 71% of demand is satisfied, while in the existing approach [15] this value is 65.1%. It shows that our approach satisfies demand at more places as compared to the existing approach. The resource allocation based on multiple objectives using priorities of tasks and places is most effective for rescue and recovery operations.

Another parameter for evaluating the performance of resource algorithm is to find out whether the resources are utilized optimally or not. It is possible to measure this using percentage utilization of particular resource. Figure 6 represents the percentage utilization of resource from each warehouse to different disaster places. As mentioned in Tables 1 and 2 there are 7 warehouse centers and 10 places that are affected by a disaster. The sum of all percentage values is 100, that is, 100% utilization of the resource. The meaning is that, from different warehouses, all the available resources of a similar type are completely utilized at different places. From this figure, it is clear that using our proposed algorithm the resource utilization is 100% which is the most desirable. This is possible because of multiobjective based allocation. Figure 6 shows the percentage of resource allocation from each warehouse to different disaster places.

The proposed approach is compared with the existing approach of resource allocation [15] in which the same set of input is given to both the existing allocation approach and the proposed allocation approach. Our proposed approach considers more objectives and constraints as compared to the existing resource allocation approach [15]. As depicted in Figures 5 and 6, the proposed approach first serves more needing places as compared to the existing approach [15].

As mentioned in Tables 1 and 2, respectively, the demand is more than the number of resources, so it is not possible to serve all places at the same time by allocating required number of resources. Few places still need to be served. As per (3), it is a situation of overdemand for resources, so the resource scheduling is required to fulfill the demand for completing the task. As shown in Figure 5, the disaster-affected places such as {P2; P3; P5} are completely unserved places, while disaster-affected places {P1; P4; P6; P7} are partially served. Thus, at all these places, resources are to be scheduled as per the availability. For this task, our proposed resource scheduling algorithm calculates the scheduling index based on the benefit parameters using TOPSIS as discussed in the previous section. For the scenario depicted here, the benefit parameter is measured in terms of entity count ECS which is served to reduce the fatality. For example, in the case of disaster, the benefit parameter can be measured in terms of how many severely injured people can be served. Various injured people can be further classified based on their age to evaluate the benefit parameter. The reason is that serving smaller children may have much more urgency as compared to middle aged people. The other parameter considered for evaluating the cost is transportation time.

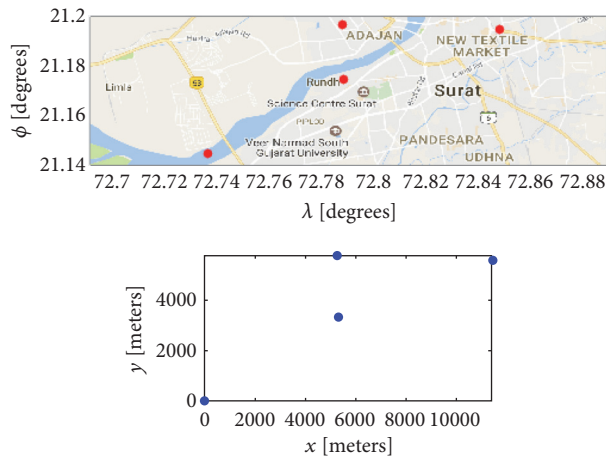


Figure 4: Distance of disaster-affected place P_4 from nearby warehouse.

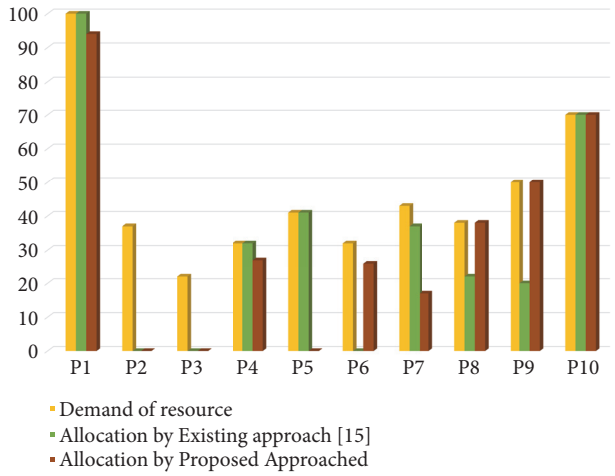


Figure 5: Comparison of resource allocation as per demand.

The third parameter which is evaluated for comparing the performance of the proposed algorithm with the existing one is overall completion time of tasks at different places altogether. In the case of postdisaster management, overall completion time of a particular activity should be minimized to reduce a loss. For evaluating overall completion time all the nonserved places that need services for a particular task should be optimally scheduled. The overall completion time is compared using our proposed algorithm and the existing approach [31]. In case of our proposed algorithm, these places are scheduled using the *Scheduling_index* which is calculated using the TOPSIS technique based on various parameters as discussed in Algorithm 2.

As mentioned earlier, the benefit parameter is evaluated in terms of entity count which is to be served for reducing the fatality. In case of disaster, the benefit parameter is evaluated in terms of how many severely injured people can be served. Various injured people can be further classified into groups based on their age. The reason is that serving smaller children may have much more urgency as compared to a middle aged group of people. Similarly, the other parameter considered for evaluating is the cost parameter which is transportation time and cost. The scheduling of resources for nonallocated disaster places are performed based on *Scheduling_index* using the available number of resources. The scheduling of the service at these places is represented using Gantt chart as depicted in Figure 7. The detail description is as follows.

It is assumed that after the task is completed at any disaster place, the resources allocated to that place are returned to the respective warehouse center. For evaluating the robustness of our algorithm, we consider a typical scenario where there is a scarcity of resources; i.e., for example, only 40% of resources are available at different warehouses in total. Further, it is also assumed that the availability of resources is measured every two hours which allows dynamic scenario for resource allocation and scheduling. From Figure 5, it is observed that the places $\{P_2, P_3, P_5\}$ are completely unserved, i.e., so far no resources are allocated at these places at the end of the resource allocation phase, so these places need to be served first. For showing the efficacy of our proposed algorithm, timeline chart is presented in Figure 7 depicting the starting time and ending time of an activity or task at a

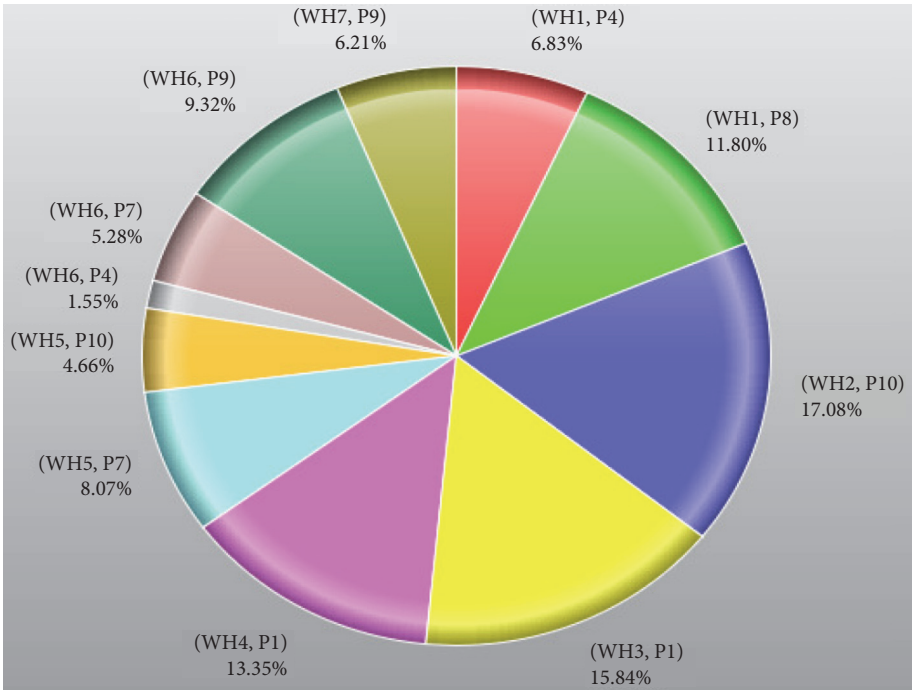


Figure 6: Percentage of resource allocation from warehouse to affected places.

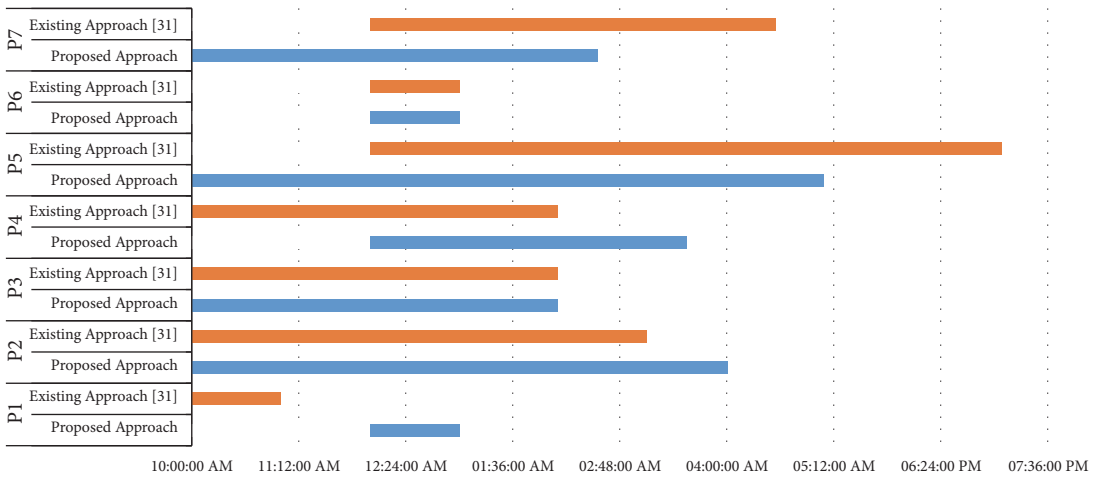


Figure 7: Gantt chart showing scheduling of nonallocated disaster places.

particular place. This represents the scheduling of resources; indirectly, it is the scheduling of activities at different places.

Let the rescue operation start at $10:00AM$ in the morning; at that instance the currently available number of resources at warehouse center is 129, which is 40% of the total resources available at warehouse center at the beginning of scheduling. Based on scheduling index calculated using our proposed Algorithm 2, the places $\{P_5, P_2, P_3, P_7\}$ need to be served first as they are more important places compared to places $\{P_6, P_4, P_1\}$. As shown in Figure 7, places $\{P_2, P_3, P_5, P_7\}$ are scheduled first at $10:00:00AM$, while places $\{P_1, P_4, P_6\}$ are served in the next scheduling phase. Using the existing approach [31], various activities are allocated with resources sequentially and, hence, the activities at these places are scheduled in sequence only, that is, P_1 to P_4, based on the availability of the resources. The existing scheduling approach [31] uses Bankers algorithm, while the proposed scheduling algorithm provides a rank to each unserved place based on multiple parameters using TOPSIS technique. It schedules resources considering availability of the resource and rank of unserved places, i.e., calculated using TOPSIS technique. At the same time when both algorithms perform scheduling, it is found that our proposed approach performs scheduling of all nonallocated places in 7 hours and 5 minutes, while

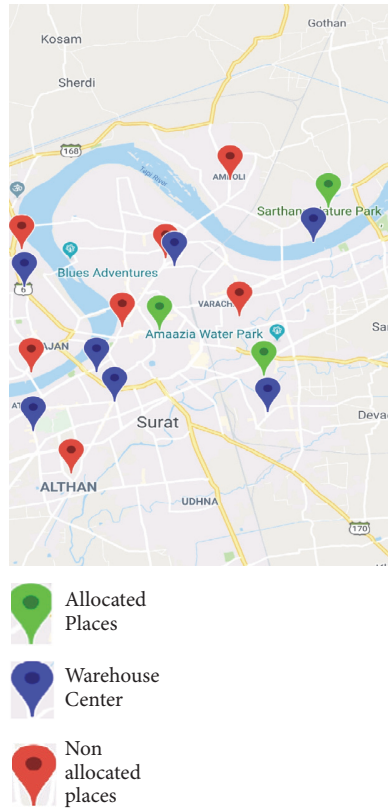


Figure 8: Visualization on Google Maps of disaster-affected places and warehouse center.

the existing approach [31] takes 9 hours and 5 minutes, i.e., the existing approach takes 2 hours more for completing the tasks.

The proposed method is also visualized using the Android application developed. The number of warehouses centers and the number of allocated places and nonallocated places are displayed using blue, green, and red color markers, respectively, using Google Maps in Figure 8. This figure represents the outcome of resource allocation phase and further scheduling is performed to all the nonallocated places.

5. Conclusion

In this paper, multiobjective based resource allocation and scheduling problems are formulated considering various parameters. These parameters allow one to assign the priority to different tasks and places depending upon the urgency of rescue and recovery operations. Using the proposed algorithm, the resources are allocated fairly at different places, in case of the demand is being more than required resources then resource scheduling is performed. Proposed resource scheduling algorithm performs successfully in terms of achieving maximal resource utilization and fulfilling the demand at various places for different tasks. The proposed algorithm is compared with the existing approach and the simulation results show that the proposed resource management approach produces better results. The current work can be extended for other parameters as per the need of different applications and the type of disaster. Also, the evaluation can be performed with more scenarios.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work is supported and funded by the Ministry of Electronics and Information Technology (MeitY), Government of India (Grant No. 13(4)/2016-CC&BT).

References

[1] "Data collection survey for disaster prevention in india," Available: http://open_jicareport.jica.go.jp/pdf/12245155.pdf, (2015).

[2] "Gujarat earthquake rehabilitation and reconstruction policy," Available: http://gsdma.org/uploads/Assets/iec/earthquakerr06172017024901390.pdf, 2001.

[3] N. Altay and W. G. I. Green III, "OR/MS research in disaster operations management," *European Journal of Operational Research*, vol. 175, no. 1, pp. 475–493, 2006.

[4] L. Pearce, "Disaster management and community planning, and public participation: how to achieve sustainable hazard mitigation," *Natural Hazards*, vol. 28, no. 2-3, pp. 211–228, 2003.

[5] D. G. Kovács and K. M. Spens, "Humanitarian logistics in disaster relief operations," *International Journal of Physical Distribution & Logistics Management*, vol. 37, no. 2, pp. 99–114, 2007.

[6] "Surat muncipal corporation (smc)," https://mysurat.in/zones .ht, 2018.

[7] S. P. Simonovic, *Systems Approach to Management of Disasters: Methods and Applications*, John Wiley & Sons, 2011.

[8] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): a vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[9] M. T. Ali, A. H. Mahmood, and T. Islam, "RFID-Based Disaster Management with Object Detection and Positioning System," in *Proceedings of the 2015 IEEE International Conference on Telecommunications and Photonics (ICTP)*, pp. 1–4, IEEE, 2015.

[10] S. Mi, X. Lee, Z. Shao et al., "LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex," *Nature Neuroscience*, vol. 7, no. 3, pp. 221–228, 2004.

[11] A. Zerger and D. I. Smith, "Impediments to using GIS for real-time disaster decision support," *Computers, Environment and Urban Systems*, vol. 27, no. 2, pp. 123–141, 2003.

[12] K. Uno and K. Kashiyama, "Development of simulation system for the disaster evacuation based on multi-agent model using GIS," *Tsinghua Science and Technology*, vol. 13, no. 1, pp. 348–353, 2008.

[13] T. Joseph Tanzi, M. Chandra, J. Isnard, D. Camara, O. Sebastien, and F. Harivelo, ""Towards" drone-borne" disaster management: future application scenarios," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, p. 181, 2016.

[14] A. Restas, "Drone applications for supporting disaster management," *World Journal of Engineering and Technology*, vol. 3, no. 03, p. 316, 2015.

[15] R. Kondaveti and A. Ganz, "Decision support system for resource allocation in disaster management," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3425–3428, IEEE, 2009.

[16] S. Bhattacharjee, S. Roy, and S. DasBit, "DPDRM: A decentralized post-disaster resource management scheme using energy efficient smart phone based DTN," *Journal of Network and Computer Applications*, vol. 111, pp. 1–16, 2018.

[17] E. Aktaş, F. Ülengin, and Ş. Önsel Şahin, "A decision support system to improve the efficiency of resource allocation in healthcare management," *Socio-Economic Planning Sciences*, vol. 41, no. 2, pp. 130–146, 2007.

[18] J. T. Blake and M. W. Carter, "A goal programming approach to strategic resource allocation in acute care hospitals," *European Journal of Operational Research*, vol. 140, no. 3, pp. 541–561, 2002.

[19] A. Athanassopoulos and C. Gounaris, "Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications," *European Journal of Operational Research*, vol. 133, no. 2, pp. 416–431, 2001.

[20] Z. Wang, W. Xu, J. Yang, and J. Peng, "A game theoretic approach for resource allocation based on ant colony optimization in emergency management," in *Proceedings of the International Conference on Information Engineering and Computer Science (ICIECS)*, pp. 1–4, IEEE, 2009.

[21] R. Kaewpuang, D. Niyato, P. Wang, and E. Hossain, "A framework for cooperative resource management in mobile cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 2685–2700, 2013.

[22] A. J. Younge and Z. Wang, "Efficient resource management for cloud computing environments," in *Proceedings of the Green Computing International Conference*, pp. 357–364, 2010.

[23] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 755–768, 2012.

[24] A. M. Caunhye, X. Nie, and S. Pokharel, "Optimization models in emergency logistics: a literature review," *Socio-Economic Planning Sciences*, vol. 46, no. 1, pp. 4–13, 2012.

[25] M. Muaafa, A. L. Concho, and J. Ramirez-Marquez, "Emergency resource allocation for disaster response: an evolutionary approach," in *Proceedings of the 12th International Probabilistic Safety Assessment and Management Conference, (PSAM '14)*, vol. 40, pp. 1–10, June 2014.

[26] F. Fiedrich, F. Gehbauer, and U. Rickers, "Optimized resource allocation for emergency response after earthquake disasters," *Safety Science*, vol. 35, no. 1–3, pp. 41–57, 2000.

[27] W. Zhao, K. Ramamritham, and J. A. Stankovic, "Preemptive Scheduling Under Time and Resource Constraints," *IEEE Transactions on Computers*, vol. C-36, no. 8, pp. 949–960, 1987.

[28] K. Jeffay, D. F. Stanat, and C. U. Martel, "On non-preemptive scheduling of period and sporadic tasks," in *Proceedings of the Twelfth Real-Time Systems Symposium*, pp. 129–139, IEEE, 1991.

[29] V. Van Peteghem and M. Vanhoucke, "A genetic algorithm for the preemptive and non-preemptive multi-mode resource-constrained project scheduling problem," *European Journal of Operational Research*, vol. 201, no. 2, pp. 409–418, 2010.

[30] J. S. Kumar, M. A. Zaveri, and M. Choksi, "Activity based resource allocation in iot for disaster management," in *Proceedings of the International Conference on Future Internet Technologies and Trends*, vol. 220, pp. 215–224, Springer International Publishing, 2017.

[31] J. S. Kumar, M. A. Zaveri, and M. Choksi, "Task based resource scheduling in iot environment for disaster management," *Procedia Computer Science*, vol. 115, pp. 846–852, 2017.

[32] S. Lauesen, "Job scheduling guaranteing reasonable turnaround times," *Acta Informatica*, vol. 2, no. 1, pp. 1–11, 1973.

[33] L. Kalinovcic, T. Petrovic, S. Bogdan, and V. Bobanac, "Modified Banker's algorithm for scheduling in multi-AGV systems," in *Proceedings of the 2011 IEEE International Conference on Automation Science and Engineering (CASE 2011)*, pp. 351–356, IEEE, August 2011.

[34] E. E. Ugwuanyi, S. Ghosh, M. Iqbal, and T. Dagiuklas, "Reliable resource provisioning using bankers' deadlock avoidance algorithm in mec for industrial IoT," *IEEE Access*, vol. 6, pp. 43327–43335, 2018.

[35] W. D. Cottrell, "Simplified program evaluation and review technique (PERT)," *Journal of Construction Engineering and Management*, vol. 125, no. 1, pp. 16–22, 1999.

[36] R. J. Luttman, G. L. Laffel, and S. D. Pearson, "Using pert/cpm (program evaluation and review technique/critical path method) to design and improve clinical processes," *Quality Management in Health Care*, vol. 3, no. 2, pp. 1–13, 1995.

[37] M. Behzadian, S. K. Otaghsara, M. Yazdani, and J. Ignatius, "A state-of the-art survey of TOPSIS applications," *Expert Systems with Applications*, vol. 39, no. 17, pp. 13051–13069, 2012.

[38] Y.-J. Lai, T.-Y. Liu, and C.-L. Hwang, "TOPSIS for MODM," *European Journal of Operational Research*, vol. 76, no. 3, pp. 486–500, 1994.

[39] C. Chen, "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.

Hindawi
Wireless Communications and Mobile Computing
Volume 2019, Article ID 7292075, 7 pages
https://doi.org/10.1155/2019/7292075

Research Article

Textile Multiantenna Technology and Relaying Architectures for Emergency Networks

Estefanía Crespo-Bardera,[1] **Adrián Vega Delgado**,[1] **Aarón Garrido Martín**,[2] **Alfonso Fernández-Durán**,[2] **and Matilde Sánchez-Fernández**[1]

[1]*The Signal Theory & Communications Department, Universidad Carlos III de Madrid, Spain*
[2]*Nokia Spain, María Tubau 9, 28050 Madrid, Spain*

Correspondence should be addressed to Estefanía Crespo-Bardera; ecrespo@tsc.uc3m.es

Received 2 November 2018; Accepted 22 January 2019; Published 5 February 2019

Guest Editor: Maurizio Casoni

Every year around 200 million people are affected by hazards of different nature. In most of these situations public protection and disaster relief personnel are usually the first responders to provide help. To provide differential relief coverage in these scenarios, novel communication and network functionalities are being demanded, relegating today's narrowband private radio (PMR) emergency systems to the background. These are data-support, increased coverage, broadband communication, and high reliability which will be addressed by novel communication technologies such as Long Term Evolution (LTE), LTE Advanced-pro, and future 5G. In this work we tackle two key technological solutions for future emergency communication networks such as an architecture based on relay nodes and enhanced user equipment by means of multiple-input-multiple-output (MIMO) techniques.

1. Introduction

In the last decade, a total of 6,090 disasters were reported globally, originated from both natural hazards such as climatological, geophysical, meteorological, hydrological, and biological and artificial or technological hazards, being industrial, transport, and miscellaneous [1]. This figure yields around 600 disasters per year around the globe.

This amount of disasters has implied that 771,911 people were deceased under those circumstances over the last 10 years. In addition, the figure for the people affected by these hazards is close to 200 million every year, understanding people affected as those who were injured, whose house was destroyed, or those who required immediate help and aid such as salvage, basic survival needs, and medical aids [2].

Data on the number of people affected and the typology of the disaster are very useful information for disaster awareness, for the potential planning of a response and for the definition of future strategies for lowering disaster impact. The reduction of disaster risk and disaster losses is the objective of the Sendai Framework for Disaster Risk Reduction (DRR) 2015–2030 [3], proposing a broad people-centered approach and a shift in emphasis from disaster management to disaster risk management. Many measures are to be tackled with this purpose: economic, structural, legal, social, health, cultural, educational, environmental, technological, political, and institutional. Nevertheless, even within this broad approach there is in most of the cases a first need for human intervention for emergency reaction after the disaster and this is where first responders, public protection, and disaster relief personnel or emergency services appear. Specifically, Priority 4 "Enhancing disaster preparedness for effective response and to *Build Back Better* in recovery, rehabilitation and reconstruction" [3] deals with, among others, contingency programs, resilience of infrastructure, workforce training, disaster risk, social technologies, hazard-monitoring telecommunications systems, and emergency communications infrastructure. Hence, technology and, in particular, communication and information systems are two

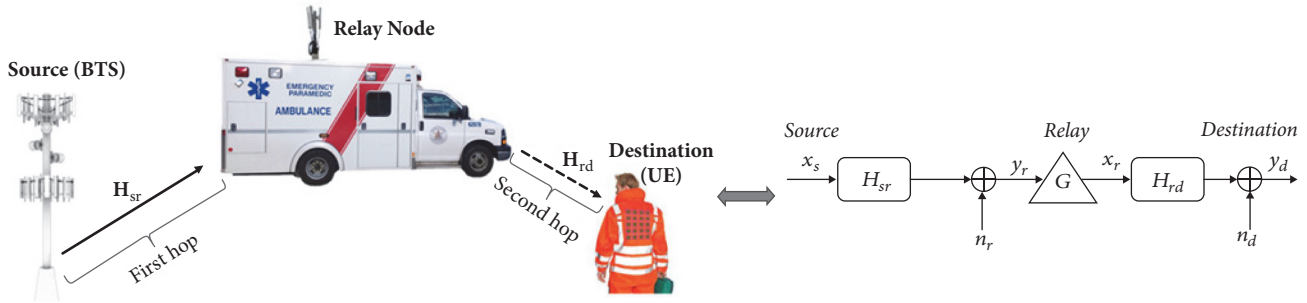


Figure 1: System model of a two-hop half-duplex single relay scenario.

of the instruments that have always played and will play a very important role on disaster managing, preparedness and intervention [4–6].

One example are social networks, they have been a distributed source of information with the ability to have local access to data in situ. Twitter and Facebook were used within the Ushahidi open-source crisis-mapping software in Haiti to gather information together with an SMS platform [7]. Also different technologies are being explored to provide first responders with enhanced communication capabilities to alleviate, resolve, or overcome an emergency situation. One of the key programs in 2016 for the US Department of Homeland Security was the usage of wearable technology that includes physiological sensors for health support, high-performance materials, communications capabilities, and context awareness.

1.1. Current and Future Public Safety and Emergency Networks. Narrowband Private Mobile Radio (PMR) systems support certain mission-critical features, i.e., device-to-device (D2D) calling, one-to-many calling, push-to-talk calling, etc. However enhanced capabilities such as robust broadband infrastructure, augmented information availability to enhance situational awareness, and rapidly deployable systems for streamline operations will rely on future wireless communication systems (LTE, LTE Advance-pro, 5G) [8] that, being aware of these advanced communications needs, are already proposing specific standardized solutions [9] for emergency scenarios.

New trends in business such as disruptive technologies, hypercompetition, new customer power and quality demands, etc. are affecting the vertical industries in different ways and this is why industries requirements are more and more demanding. Here is where 5G technologies, specifically network slicing, play a very relevant role. Traditionally, telecommunication networks were based on physical nodes that provided all the services that the operator offered and whose configuration was not performed automatically. Information and communications technologies (ICT) development have led to understand networks in a different way, introducing the concept of logical network, where networks are designed, instantiated, and operated on-demand, meeting the requirements of the business, customers, or services. These logical networks are named network slices. A network slice, hence, covers a complete set of network functions to give a specific service. Therefore, network slices can tackle a huge variety of technical aspects to meet the requirements for each specific service. These requirements can go from latency, throughput (uplink and downlink), availability and resilience, reliability or coverage, to communication security aspects [10]. Public safety is one of the vertical industries whose requirements will be met with 5G technologies whose most important needs are a reliable connectivity, coverage, and enough throughput to transfer any content that may be helpful in a disaster situation.

In this work we explore two key aspects of public safety vertical: high-throughput and connectivity/coverage. With this aim, a network architecture based on relays for a rapidly deployable communication infrastructure is presented and spectral efficiency is improved by means of multiple-input-multiple-output (MIMO) antenna technology. MIMO technology here has a different approach which is targeting the user end by seamlessly deploying a large number of antennas in textile technology.

2. Technological Solution to Address Broadband Coverage

In this section we present a network architecture proposal for future public safety communications that may improve important weakness of current PMR networks such as the limited coverage and spectral efficiency giving rise to the use of multimedia services. The architecture envisions a two-hop relay network, merging massive MIMO textile technology into the LTE cellular network.

As depicted in Figure 1, in the proposed network three main components (base transceiver station (BTS), relay node, and user equipment (UE)) and two main links can be distinguished. One of them defines the base station coverage area which is defined between the base station and a relay node and the second one establishes the relay node coverage area between the repeater device and the user equipment.

2.1. Relay Node. Thanks to the Long Term Evolution-Advanced (LTE-A) standard defined by the 3GPP, different methods to improve coverage at relatively low cost to the operator have been brought in. Among them relay nodes can be highlighted.

Relay nodes are classified into different categories attending to different characteristics such as the level of the protocol stack in which the user traffic is transmitted, and the functions enabled in the control plane [11], and the transparency [12] with respect to the user or the use of the spectrum [13]. A more detailed explanation is given below.

2.1.1. Level in the Protocol Stack. Layer 1, 2, and 3 relay nodes can be distinguished in this category. Layer 1 relays (also known as repeaters) act amplifying and forwarding the downlink and uplink signals between the user equipment (UE) and the BTS, just like a simple analog amplifier. This fact allows extending the coverage area of an existing base station to locations where it cannot reach by itself. It should be noticed that Layer 1 relays work in a nontransparent mode and typically the noise from the link between the base station and the relay is amplified in the next link. Layer 2 and 3 relays essentially decode the incoming signal and remodulate and reencode it before the amplified version is forwarded. These kinds of relays avoid drawbacks such us noise being amplified and retransmitted.

2.1.2. Spectrum Use. Two types of relays can be distinguished depending on the use of the spectrum in the different links: in-band and out-of-band. In the first case, the link between the base station and the relay shares the same frequency as the link between the relay and the UE. On the other case, the link between the BTS and the relay does not share the same frequency as the link between the relay and the UE.

2.1.3. Transparency. The main difference between nontransparent and transparent operation modes lies in how framing information is transmitted. In nontransparent mode, the relay nodes transmit frame header information mainly containing scheduling information that it is useful for the nodes to know when they can transmit and receive information. Conversely, in transparent operation mode relay nodes do not transmit frame header information.

From now on, attending to the protocol stack classification we will assume an amplify-and-forward relay scenario.

2.2. Massive MIMO Textile Technology in the UE. To enable high spectral efficiency necessities of current public safety networks, the textile antenna technology, which allows us to deploy MIMO-based personal area networks in the proximity of the human body, appears to be a promising solution [14]. The main idea behind this technology consists on embedding a large textile antenna array which acts as a transmitter/receiver at the user's garments (see Figure 2), bringing the benefits of massive MIMO directly to the user's end. It must be noticed that communications using the textile technology may be on-body (wireless communication link between devices placed in the body of the user) or off-body (wireless link is established between devices place in the body such as a video camera and an external element like the base station).

Figure 2: Real large textile antenna array deployed at user jacket backside.

Table 1: Design features of the large textile antenna using CST Microwave Studio.

Working Frequency	2.5 GHz
Planar array Size	8×5
Inter-element distance (avoiding Mutual Coupling)	0.66λ
Dielectric Substrate	Common Felt ($\epsilon_r = 1.38$)
Metallization	Electrotextile Material
Array Area	$44.6 \times 60.6\ \text{cm}^2$
Backward power levels	−20 dB
Bandwidth	70 MHz

We consider a textile planar array whose main features are described in [14, 15] and summarized in Table 1. The number of antennas of this particular design is $N = 40$; nevertheless other configurations will also be explored in Section 4.

3. MIMO-Relaying Scenario Modeling and Performance

We analyze a two-hop single relay scenario with multiple radiating elements at the source/BTS (M), relay (R), and destination/UE (N) working in an amplify-and-forward mode (Layer 1 relay). The relay node is assumed to work in half-duplex mode. This means that communication from source to relay and from relay to destination is carried out in two different time slots. Additionally, we assume that there is no direct link between the source and destination.

During the first time slot, the transmitted signal from the source propagates through the first-hop channel, and the received signal at the relay is given by

$$\mathbf{y}_{\mathrm{r}} = \sqrt{\rho_{\mathrm{r}}}\mathbf{H}_{\mathrm{sr}}\mathbf{x}_{\mathrm{s}} + \mathbf{n}_{\mathrm{r}} \tag{1}$$

where $\mathbf{x}_{\mathrm{s}}$ is the $M \times 1$ transmitted signal vector with normalized power, $\mathbf{H}_{\mathrm{sr}}$ is the $R \times M$ channel matrix between the source and the relay, ρ_{r} is the signal-to-noise ratio (SNR) of the source-relay link, and $\mathbf{n}_{\mathrm{r}}$ is the $R \times 1$ relay noise vector with normalized variance.

During the second time slot, the relay amplifies the received signal defined in (1) and forwards it to the destination. The received signal at the UE is as follows:

$$\mathbf{y}_{\mathrm{d}} = \sqrt{\rho_{\mathrm{d}}}\mathbf{H}_{\mathrm{rd}}\mathbf{G}\mathbf{H}_{\mathrm{sr}}\mathbf{x}_{\mathrm{s}} + \left[\sqrt{\rho_{\mathrm{d}}}\mathbf{H}_{\mathrm{rd}}\mathbf{G} \;\; \mathbf{I}\right]\begin{bmatrix}\mathbf{n}_{\mathrm{r}}\\ \mathbf{n}_{\mathrm{d}}\end{bmatrix} \tag{2}$$

where $\mathbf{H}_{\mathrm{rd}}$ is the $N \times R$ channel matrix between the relay and the destination, $\mathbf{G}$ is the $R \times R$ relay amplification factor, ρ_{d} is the relay-destination link SNR, and $\mathbf{n}_{\mathrm{r}}$ is the $N \times 1$ destination noise vector with normalized variance.

3.1. Channel Model. In our analysis, the first-hop channel $\mathbf{H}_{\mathrm{sr}}$ is modeled by a Rayleigh i.i.d. fading distribution, assuming a full-scattering environment and large spacing among antennas at both, the BTS and the relay. The second hop channel $\mathbf{H}_{\mathrm{rd}}$, defined between the UE equipped with MIMO textile technology and the relay node, is modeled to take into account the textile antenna parameters and radiation patterns in transmission and reception and the surrounding scattering environment. Each element h_{nm} in the channel matrix $\mathbf{H}_{\mathrm{rd}}$ is described by the Green's function sampled at the position of the n-th receiving antenna $\mathbf{r}'_n$ given that the point source is located at the m-th transmitting antenna ($\mathbf{r}_m$) [16]:

$$\begin{aligned} h_{nm} = {} & \int\int \mathcal{G}_m(\theta,\phi)\,\mathcal{G}'_n(\theta',\phi')\,S\left(\mathbf{k}'(\theta',\phi'),\mathbf{k}(\theta,\phi)\right) \\ & \cdot e^{-j\mathbf{k}(\theta,\phi)\mathbf{r}_m} e^{j\mathbf{k}'(\theta',\phi')\mathbf{r}'_n} d\mathbf{k}'(\theta',\phi')\, d\mathbf{k}(\theta,\phi). \end{aligned} \tag{3}$$

where $\mathcal{G}_m(\theta,\phi)$ and $\mathcal{G}'_n(\theta',\phi')$ represent the radiation patterns in azimuth (ϕ) and elevation (θ) at the transmitter and receiver, respectively, $\mathbf{k}(\theta,\phi)$ and $\mathbf{k}'(\theta',\phi')$ are the wave vector space at the transmitter and receiver, respectively, and $S(\mathbf{k}'(\theta',\phi'),\mathbf{k}(\theta,\phi))$ is the channel scattering function, which relates the plane wave's emitting and receiving directions, $\mathbf{k}$ and $\mathbf{k}'$, respectively. It must be noticed that this channel model assumes single-polarized antennas as the ones described in Section 2.2 and therefore the transmit and receive vector fields together with the channel scattering function can be considered scalars. However, this model could be easily extended to further consider different polarization if we work with vectorial fields and a dyad transformation function as scattering [17, 18].

The vector space can be sampled into L plane waves at the transmitter and L' plane waves at the receiver to cover the entire space. Then, the channel matrix can be decomposed [19] into the product of five matrices $\mathbf{H}_{\mathrm{rd}} = \mathbf{B}^{\dagger}_{rx}\Sigma_{rx}\mathbf{H}_{\mathrm{iid}}\Sigma_{tx}\mathbf{B}_{tx}$, where $\mathbf{B}_{rx}$ ($L' \times N$) and $\mathbf{B}_{tx}$ ($L \times R$) are beamforming matrices depending on the antenna geometry and radiation patterns, $\mathbf{H}_{\mathrm{iid}}$ ($L' \times L$) is a complex Gaussian random matrix with i.i.d. components and variance one, and Σ_{rx} and Σ_{tx} are normalized diagonal matrices whose main diagonal is shaped with the corresponding angular power spectra. The MC between antenna elements is then captured via the coupling matrices $\mathbf{C}_{tx}$ and $\mathbf{C}_{rx}$ as shown in [12]. The resulting channel matrix is given by $\mathbf{H}_{\mathrm{rd}} = \mathbf{C}_{rx}\mathbf{B}^{\dagger}_{rx}\sum^{1/2}_{rx}\mathbf{G}\sum^{1/2}_{tx}\mathbf{B}_{tx}\mathbf{C}_{tx}$.

3.2. Relay Amplification Factor. We assume that the relay node simply applies a linear transformation on the received signal $\mathbf{y}_{\mathrm{r}}$ by means of the relay amplification factor $\mathbf{G}$ as follows:

$$\mathbf{G} = \beta\mathbf{I}_R \tag{4}$$

Parameter β is defined in order to achieve the relay-destination link SNR ρ_{d} [20, 21]:

$$\beta = \sqrt{\frac{1}{\rho_{\mathrm{r}}\left\|\mathbf{H}_{\mathrm{sr}}\right\|^2_F + 1}} \tag{5}$$

where $\|\cdot\|^2_F$ represents the squared Frobenius norm. It must be noticed that in our model the power is assumed to be allocated uniformly over all the antennas.

3.3. Achievable Rates for the Relay Scenario. The achievable rates of the amplify-and-forward half-duplex MIMO-based two-hop single relay channel with $\mathbf{G}$ as in (4) and (5) can be written as in (6) [22, 23].

Since the transmission is completed in two different time instances, the spectral efficiency is reduced by half and the SNR of each of the links are doubled.

4. Results and Discussion

In this section, we evaluate the performance of the two-hop relay architecture in terms of achievable rates with different antenna configurations. On the one hand, we analyze the achieved capacity in one scenario where the textile array deployed at the UE is built with $N = 40$ antennas and the BTS and relay count with $M = R = \{1, 4, 8, 20\}$. On the other hand, we analyze the optimal relay position to maximize the achieved capacity for different antenna configurations. In this particular case, we consider the same number of antennas in all communications ends $M = R = N = \{1, 2, 4, 6, 10\}$ and study the impact of the SNR level on the optimization. Simulations are carried out generating 500 channel samples using the channel model presented in Section 3.

Figure 3 shows the achieved capacity when different levels of SNR. At low SNR values such as SNR = 0 dB data rates values from approximately 1.6 Mbps up to 48 Mbps can be achieved considering a bandwidth of 10 MHz. On the other hand, at high SNR regimes like SNR = 10 dB, achieved data rates go from 10 Mbps up to 200 Mbps. As expected, the achieved capacity increases as the number of antennas increases.

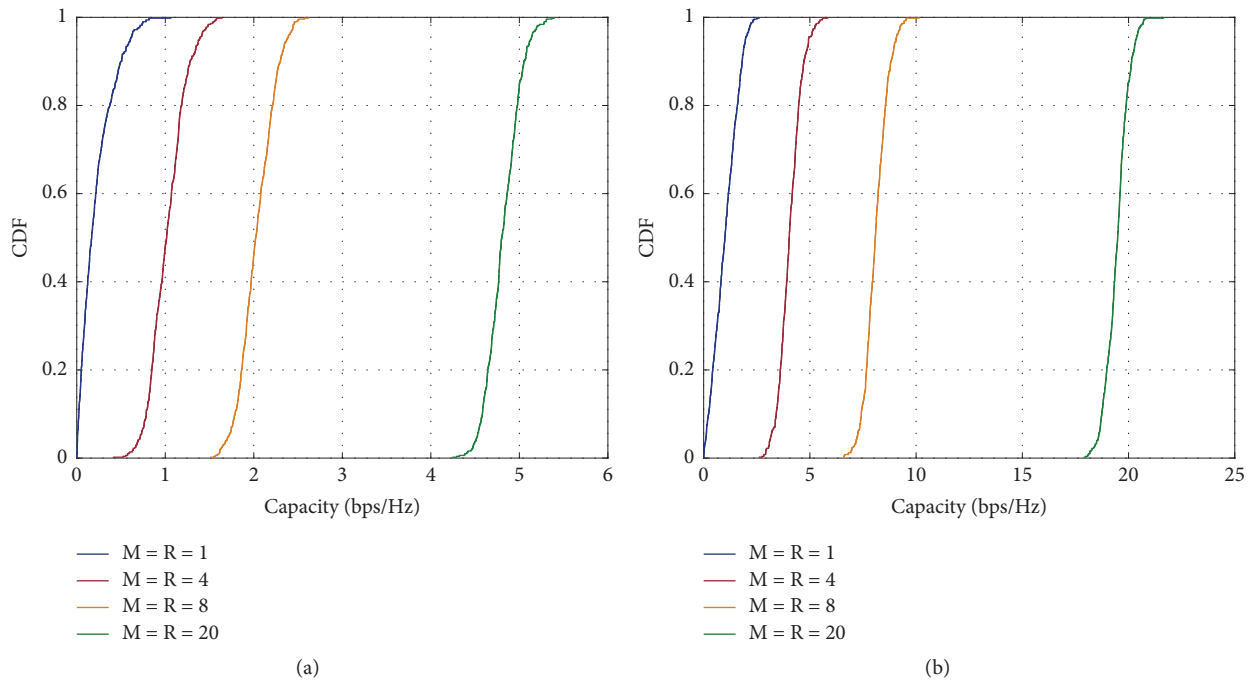


FIGURE 3: Achievable rates with different antenna configurations (a) achieved rates for $N = 40$, $R = M = \{1, 4, 8, 20\}$, and SNR(dB) = 0; (b) achieved rates for $N = 40$, $M = R = \{1, 4, 8, 20\}$, and SNR(dB) = 10.

4.1. Optimal Relay Allocation. Scenarios with low SNR are the candidates for the deployment of a MIMO-based relay solution to increase coverage or improve rates. To illustrate that, we will show next the average achievable rates in different positions of the relay. The SNR model assumes an exponent path loss of 3.32 which has been measured in outdoor-to-indoor scenarios with the UE equipped with textile antennas placed underground [24]. The plots include as a reference the achievable rates of a direct link scenario, with no relay, showing which is the best relay configuration (normalized position) to improve the direct link achievable rates.

$$R = \frac{1}{2}\log_2 \det\left(\mathbf{I}_N + \frac{4\rho_d\rho_r}{M}\frac{\mathbf{H}_{rd}\mathbf{H}_{sr}\mathbf{H}_{sr}^{\dagger}\mathbf{H}_{rd}^{\dagger}}{2\rho_d\mathbf{H}_{rd}\mathbf{H}_{rd}^{\dagger}+2\rho_r\left\|\mathbf{H}_{sr}\right\|_F^2\mathbf{I}_N+\mathbf{I}_N}\right) \quad (6)$$

From Figure 4 it is possible to corroborate that the relay node presence appears to be especially useful at low SNR scenarios. In fact, as the SNR value increases, the performance of the relay network in terms of capacity falls slightly below the direct link case. Depending on the existing number of antennas, the coverage region where the relay node is useful (relay capacity is higher than the direct link capacity) varies, being smaller as the number of antennas increases. Furthermore, from the results in Figure 4 we can also obtain the optimal relay position leading to the maximum achievable rate in those scenarios where the relay architecture outperforms the direct link transmission.

5. Conclusions

This work proposes a network and physical layer solution based on LTE-A and future 5G capabilities to improve public safety communications, which are currently conveyed through narrowband PMR systems and mainly focused on offering limited voice services. We have analyzed the performance in terms of capacity of an amplify-and-forward relay network when massive MIMO textile technology is deployed at the user side. Additionally, we have evaluated the optimal relay location with the purpose of maximizing the achieved capacity in the two-hop network. Simulation results illustrate the viability of the proposed design, specifically for low SNR scenarios where the relay node will allow us to extend the coverage and the MIMO textile technology to improve the capacity.

Data Availability

We have no included a data availability statement in our article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

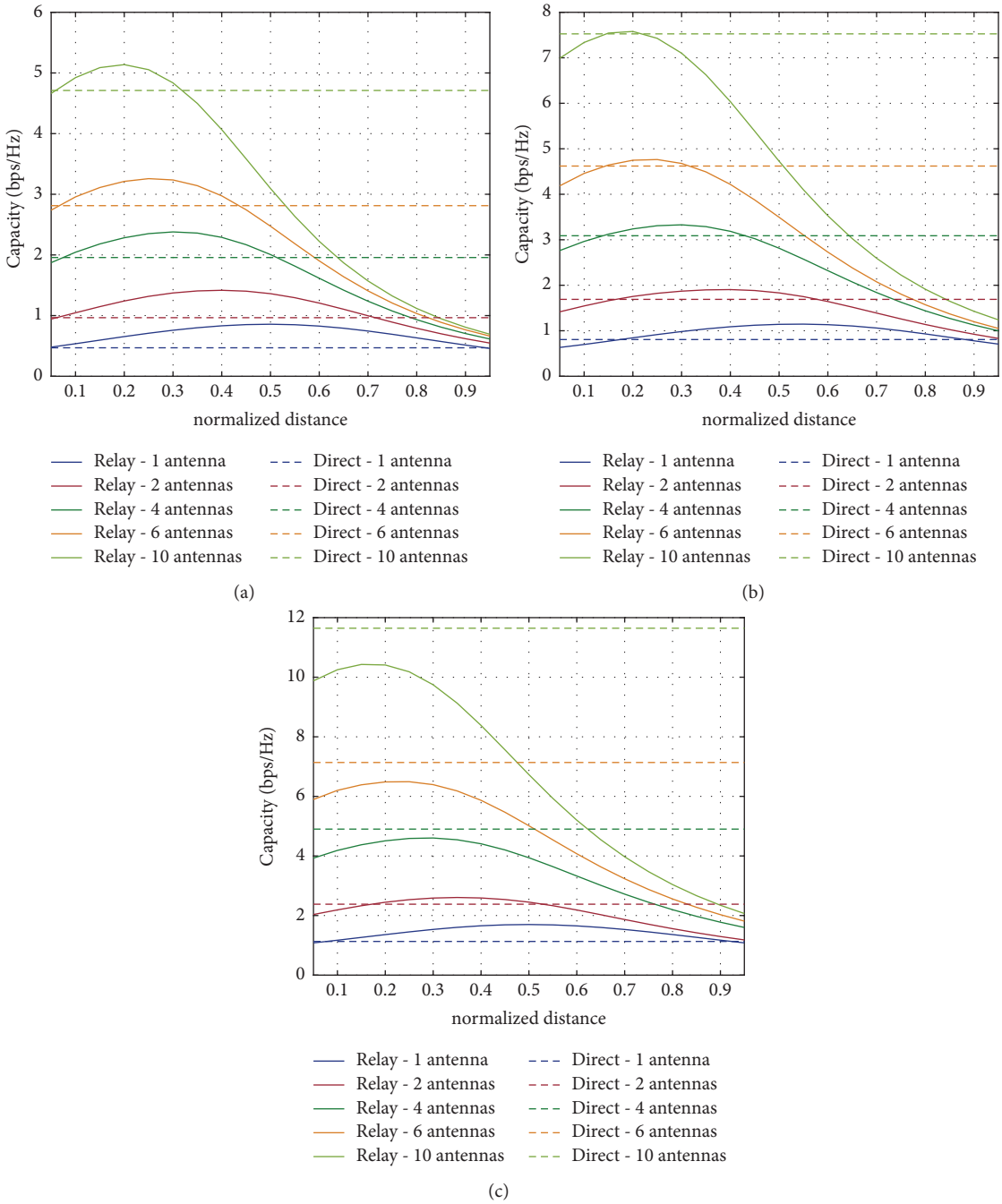


FIGURE 4: Relation between the normalized relay location and the achieved capacity at different SNR regimes and antenna configurations, considering $M = R = N$. (a) $SNR = -3$ dB; (b) $SNR = 0$ dB; and (b) $SNR = 3$ dB.

Acknowledgments

This work has been partly funded by the Spanish Government through Projects CIES (RTC-2015-4213-7), MIMO-TEX (TEC2014-61776-EXP), and TERESA-ADA (TEC2017-90093-C3-2-R) (MINECO/AEI/FEDER, UE).

References

[1] D. Sanderson and A. Sharma, *World Disasters Report, Resilience: Saving Lives Today, Investing for Tomorrow*, International Federation of Red Cross and Red Crescent Societies, 2016.

[2] D. Guha-Sapir and R. Below, "Collecting data on disasters: easier said than done," *Asian Disaster Management News*, vol. 12, no. 2, pp. 9-10, 2006.

[3] United Nations Office for Disaster Risk Reduction, "Sendai framework for disaster risk reduction 2015 - 2030," Tech. Rep., Sendai, Japan, 2015, https://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf.

[4] I. Kamen, "A new approach to disaster communication and control systems," *Electrical Engineering*, vol. 81, pp. 535–541, 1962.

[5] T. C. Chan, J. Killeen, W. Griswold, and L. Lenert, "Information technology and emergency medical care during disasters," *Academic Emergency Medicine*, vol. 11, no. 11, pp. 1229–1236, 2004.

[6] R. Shaw, T. Izumi, and P. Shi, "Perspectives of science and technology in disaster risk reduction of asia," *International Journal of Disaster Risk Science*, vol. 7, no. 4, pp. 329–342, 2016.

[7] J. Heinzelman and C. Waters, "Crowdsourcing crisis information in disaster-affected Haiti," Tech. Rep., United States Institute of Peace, 2010.

[8] T. Doumi, M. F. Dolan, S. Tatesh et al., "LTE for public safety networks," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 106–112, 2013.

[9] 3GPP, "Isolated evolved universal terrestrial radio access network (E-UTRAN) operation for public safety, stage 1," Tech. Rep. TS 22.346, Release 13, 2014.

[10] GSA, *5G network slicing for vertical industries*, GSA White Paper, 2017.

[11] M. Iwamura, H. Takahashi, and S. Nagata, "Relay technology in LTE-advanced," *NTT DoCoMo Technical Journal*, vol. 12, no. 2, pp. 29–36, 2010.

[12] H. Holma and A. Toskala, *LTE advanced: 3GPP solution for IMT-Advanced*, John Wiley and Sons, 2012.

[13] 3GPP, "Feasibility study for further advancements for E-UTRA (LTE-advanced)," Tech. Rep. G. TR 36.912, 2012.

[14] M. Sanchez-Fernandez, A. Tulino, E. Rajo-Iglesias, J. Llorca, and A. G. Armada, "Blended antenna wearables for an unconstrained mobile experience," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 160–168, 2017.

[15] E. Crespo-Bardera, M. Sánchez-Fernández, A. García-Armada, A. G. Martín, and A. F. Durán, "Analysis of a LTE-based textile massive MIMO proposal for public safety networks," in *Proceedings of the 86th IEEE Vehicular Technology Conference, VTC Fall '17*, pp. 1–5, September 2017.

[16] D. Chizhik, "Slowing the time-fluctuating MIMO channel by beam forming," *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1554–1565, 2004.

[17] T. Svantesson, "A physical MIMO radio channel model for multi-element multi-polarized antenna systems," in *Proceedings of the IEEE 54th Vehicular Technology Conference, VTC Fall '01*, vol. 2, pp. 1083–1087, IEEE, October 2001.

[18] M. L. Pablo-González, M. Sánchez-Fernández, and E. Rajo-Iglesias, "Combination of the three types of diversity to design high-capacity compact MIMO terminals," *IEEE Antennas and Wireless Propagation Letters*, vol. 13, pp. 1309–1312, 2014.

[19] P. N. Fletcher, M. Dean, and A. R. Nix, "Mutual coupling in multi-element array antennas and its influence on MIMO channel capacity," *IEEE Electronics Letters*, vol. 39, no. 4, pp. 342–344, 2003.

[20] L. Jiang, L. Thiele, and V. Jungnickel, "Modeling and measurement of MIMO relay channels," in *Proceedings of the IEEE 67th Vehicular Technology Conference-Spring, VTC '08*, pp. 419–423, IEEE, May 2008.

[21] T. Q. Duong, L.-N. Hoang, and V. N. Q. Bao, "On the performance of two-way amplify-and-forward relay networks," *IEICE Transactions on Communications*, vol. E92-B, no. 12, pp. 3957–3959, 2009.

[22] P. Herhold, E. Zimmermann, and G. Fettweis, "On the performance of cooperative amplify-and-forward relay networks," *ITG FACHBERICHT*, vol. 2, pp. 451–458, 2004.

[23] M. Herdin, "MIMO amplify-and-forward relaying in correlated MIMO channels," in *Proceedings of the 5th International Conference on Information, Communications and Signal Processing*, pp. 796–800, IEEE, 2005.

[24] E. Crespo-Bardera, M. Rodríguez, M. Sánchez-Fernández et al., "Empirical rates characterization of wearable multi-antenna terminals for first-responders," *IEEE Access*, vol. 7, pp. 6990–7000, 2019.