IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 9, 2017, accepted October 25, 2017, date of publication November 1, 2017, date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2768800

Heterogeneous Public Safety Network Architecture Based on RAN Slicing

DANIA MARABISSI, (Senior Member, IEEE), AND ROMANO FANTACCI, (Fellow, IEEE)
Department of Information Engineering, University of Florence, 50121 Firenze, Italy

Corresponding author: Dania Marabissi (dania.marabissi@unifi.it)

This work was partially supported by the Project GAUChO-A Green Adaptive Fog Computing and Networking Architecture through the MIUR-PRIN 2015 under Grant 2015YPXH4W_004.

ABSTRACT Efficient communications are of paramount importance to improve public safety (PS) operations allowing better coordination, higher situation awareness, lower response times, and higher efficiency during emergency. Consequently, the evolution of PS communication networks toward commercial broadband networks is widely well accepted. However, this evolution has to cope with several challenges, such as the provision of sufficient communication capacity, coverage, and resilience as well as deployment costs and efficient exploitation of radio resources. This has triggered the need of new architectural solutions. In this paper, we propose a heterogeneous network communication architecture where both infrastructures and spectrum are shared between PS and commercial operators thus reducing deployment costs and times, and addressing the main challenges of PS communications. The shared radio access network (RAN) is managed by means of network slicing and resources virtualization. The proposed architecture is based on a three-tier scheduler that allows to manage different network layers and different RAN slices. Numerical results derived by means computer simulations are provided in order to highlight the efficiency and flexibility of the proposed architecture in comparison with benchmark alternatives.

INDEX TERMS Public safety, network architecture, scheduling, RAN slicing, heterogeneous networks, spectrum sharing.

I. INTRODUCTION

It is widely recognized that efficient and reliable communications are essential in any activity related to a Public Safety (PS) situation. The capability of PS operators to communicate each other and share critical information using a wide range of data-centric services, gives better situational awareness and quicker response time, and speeds up the emergency management. Current PS communication networks, such as terrestrial trunked radio (TETRA), TETRA for police (TETRAPOL), or Association of Public-Safety Communications Officials-Project 25 (APCO P25), have been designed to provide a rich set of voice-oriented services, advanced security features and specific functionalities, but the support of advanced high data rate services is still lagging behind that provided by broadband commercial mobile networks. As a consequence, there is an increasing interest of governments, PS operators and research communities in improving the capabilities of PS communication systems. The envisaged solution is to provide optimized and reliable services for professional users by exploiting new technologies such as LTE (Long Term Evolution) and its foreseen 5G successor [1]. The use of a common technology for both commercial and PS communications enables synergies and offers new opportunities. In addition, PS operators can always benefit from an up-to-date communication system thus avoiding new technological gaps. Nevertheless, this evolutionary trend needs to take into account some issues. In particular, commercial networks have not been designed to guarantee all the essential attributes of PS communications. These need resilient and largely available infrastructures, able to provide a number of primary features for the connection of groups of people. In particular, PS networks must be able to establish quick and reliable individual and group voice calls, use prioritisation and pre-emption mechanisms for congestion and emergency management. Furthermore, they have to enable direct mode operation between terminals and guarantee communication security through high-level voice encryption.

Since Release 12 the 3rd Generation Partnership Project (3GPP) is working to incorporate all essential PS attributes in LTE standard in order to emulate current PS systems. In particular, LTE Release 12 has mainly focused on proximity services (ProSe), mission critical push-to-talk (MCPTT) and group communications [2], [3] that have been improved in Release 13 together with the introduction of other enhanced services such as relay communications and isolated Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) operations (i.e., a base station -BS- operating without backhaul connections). In addition to these basic services and functionalities, PS networks must be reliable, resilient and always available. Hence, future PS networks have to face with other major challenges related to network coverage, network congestion, ubiquitous connectivity and resource availability that require advanced technical and architectural solutions.

This paper proposes a new communication architecture based on the integration of a shared radio access network (RAN) with a dedicated one, thus limiting costs and time of network deployment while guaranteeing PS requirements. The shared RAN is managed by means network slicing and resource virtualization. In particular, a three-tier resource scheduler is proposed in order to have an efficient management of the radio resources in both segments of RAN.

The paper is organized as follows. After a review of the related literature presented in Sect. II, a brief description of the major challenges that PS networks have to face with is provided in Sect. III. Then the proposed architecture is described in Sect. IV with a focus on RAN slicing and radio resource management, and its performance is evaluated in Sect.V. Finally conclusions are drawn in Sect.VI.

II. RELATED LITERATURE

PS communication networks demand high quality services with continuity, timeliness and reliability as these services are related to safety-critical operations with high risk of loss. Moreover, PS networks must accomodate the data traffic peak that occurs in case of an emergency. With the aim to meet all these constraints different architectural and technological solutions have been recently proposed [4].

Device-to-Device (D2D), also known as *ProSe*, is considered a key technology to offload traffic from the network, to extend the coverage and to allow communications even when the network infrastructure is unavailable [2]. With the goal of integrating Wireless Sensor Networks (WSNs) in a PS communication system, Usman *et al.* [5] propose a software defined network (SDN) architecture supporting hierarchical D2D communications. A centralized SDN controller communicates with the mobile-cloud-heads thus reducing the number of LTE communication links and the energy consumption. Similarly, in [6] the performance of a PS network architecture based on D2D and relay-assisted transmissions, is evaluated in terms of capacity and power saving.

Network capacity and coverage can be improved by using *small cells* to create multiple access layers according to the Heterogeneous Networks (HetNets) paradigm. Small cells can be placed in critical areas where additional capacity is required by PS operators or can be temporarily deployed where the network infrastructure is completely or partially destroyed. For example, in [7], the coverage offered by the regular LTE network is enhanced by using temporary cognitive femto cells and suitable interference management strategies. Similarly, small cells are opportunistically deployed in order to provide dedicated access to PS users in critical areas in [8]. These small cells share resources with macrocells that provide basic network coverage. Also the use of *moving cells* is gaining great interest. In [9] network mobility is presented as a means to meet PS communication requirements in an isolated E-UTRAN operation. Likwise, the performance of out-of-coverage in terms of throughput is evaluated in [10] when mobile cells are used. A mobile BS-based architecture able to adapt to the traffic demand is proposed in [11]. The PS network is composed by sparsely placed BSs (in fixed positions) for supporting light traffic, and a set of mobile BSs deployed ad-hoc in emergency areas. Terrestrial communication infrastructure can be also complemented with small cells from the sky. *Unmanned aerial vehicles (UAV)* can be equipped with communication hardware and sent to suitable positions in the emergency area to augment the operation of PS networks. In [12] a network architecture exploiting UAV is proposed, and the throughput gain obtained by exploiting the mobility feature of UAVs is analysed.

Another important issue, widely investigated in the literature, is related to the lack of dedicated wideband spectrum for PS networks that significantly limits the provision of high data rate services. In reference to this, *spectrum sharing* has emerged as a potential solution [13], [14] when managed and supported by suitable algorithms and radio architectures. [15], [16] present algorithms to suitably allocate radio resources to commercial and PS User Equipments (UEs) belonging to the same RAN. The algorithms are based on priority [15] and network pricing and call admission control [16] policies to guarantee the required quality of service to each class of UE and service. Differently, in [17] two RANs are available for PS and commercial UEs (C-UEs), however, PS-UEs can also access to the commercial RAN, and their traffic is redirected by the core network. Moreover, in addition to a dedicated bandwidth, PS communications can be allocated also in the shared commercial bandwidth portion in case of emergency, and served with the highest priority resorting to LTE retention priority mechanism. The lack of radio resources during emergency situations can be also solved by means cognitive radio approaches as outlined in [18] and [19].

Differently from previous works, this paper proposes a new architecture that merges *HetNet* and *spectrum sharing* paradigms. Dedicated small cells integrate the basic commercial RAN in order to improve PS network converge and capacity, where and when needed. Dedicated small cells and cells of basic commercial network share radio resources.

This two-component-RAN is managed by means suitable resource managers and introducing RAN slicing concept to achieve spectrum efficiency and at the same time to address PS networks challenges.

III. MAJOR CHALLENGES IN FUTURE PS NETWORKS

The increasing demand of data-centric services from PS users joint with the need of reducing the technological gap with the commercial world, are making current PS paradigms based on *dedicated infrastructures*, *spectrum* and *technology* not longer suitable for future networks. This has led a general consensus toward a unique technology for PS and commercial networks. However, the issues regarding infrastructures and radio resources sharing are still open. In particular, possible architectural solutions have to cope with several challenges:

- **Network deployment** - It is straightforward to understand that the best solution for PS users is a PS-dedicated network infrastructure. This has several advantages in terms of network optimization, management and security. However, long deployment times and the need of huge investments represent critical drawbacks of this solution. On the other end, providing PS services over an already deployed commercial network is a cost-efficient and quick solution, but this would make all the sensitive network features not directly controlled by PS network operators, with the risk of loosing critical PS constraints. Hence, a suitable trade-off seems to be resorting to hybrid solutions, based on a partial infrastructure sharing, thus having a significant reduction of costs with the fulfillig of all PS requirements. Moreover, hybrid solutions could allow higher flexibility in terms of spectrum usage, service management, radio access policies, and territory coverage [20].
- **Network congestion** - In case of an emergency, networks activity increases causing traffic congestion. This is particularly critical when PS network shares infrastructures and radio resources with commercial network: PS communications compete with general public communications with unpredictable effects on high-priority emergency services. Hence, it is of paramount importance to find solutions that prevent as much as possible network congestion, and are able to guarantee critical communications when network congestion occurs.
- **Network coverage: ubiquitous connection and high throughput** - Differently from commercial operators that are mainly interested in offering services in high revenue regions (accepting the presence of uncovered areas), PS operators needs a base of granted services always available without the risk of jeopardized coverage. However, LTE provides less coverage area when compared to current PS systems with consequent low data rate and dropped communications at the cell-edge. This requires higher cell density and, hence, higher deployment costs. Consequently, coverage holes must be suitably managed using smart solutions.
- **Network resilience** - PS networks must be characterized by high reliability and resilience, also allowing communications in disaster or critical situations (e.g., earthquake, flooding, loss of power supply) when commercial networks are often seriously damaged or out of order, since in general these are not suitably designed according to the principle of redundancy typical of PS networks.
- **Radio resources** - Efficient PS broadband communications require sufficient radio resources to satisfy all the requested services. In particular, resources should be sized to meet the traffic peak when an emergency occurs, thus to avoid service unavailability. However, this approach is inefficient because leads to the waste of precious resources during routine activities (traffic pattern in PS networks significantly changes during routine or emergency activities), and it is often unfeasible due to the high costs (governments cannot earn from commercial licenses).

IV. PROPOSED ARCHITECTURE

This paper proposes a new PS network architecture that *partially shares resources and infrastructures* with commercial networks and allows to efficiently cope with the main challenges listed before. The aim of our proposal is that of designing a network that is resilient, efficient, flexible and able to overcome the commercial network limits by using HetNet concept.

A. HETNETS

HetNets represent a new networking paradigm introduced to boost capacity and coverage moving the transmitter and the receiver closer together. The basic idea is to increase cells density creating multiple access layers made of macrocells overlaid with small cells, such as microcell, picocell, femtocell and relay nodes, which are low-power and low-cost access nodes that can be deployed anywhere. This allows the system to provide high data rates, offload excess traffic from the macrocell and provide dedicated capacities to homes, enterprises, urban hot-spots or, as in our case, to PS applications and services.

B. PROPOSED INFRASTRUCTURE SHARING

The full-IP nature of LTE networks benefits and eases the success of shared network architectures. In particular, 3GPP [21] has defined two different network sharing configurations: (i) the Multi-Operator Core Network (MOCN) in which only the Radio Access Network (RAN) is shared, and (ii) the Gateway Core Network (GWCN) where also the mobility management entity (MME) and serving gateway (S-GW) are shared. For our purposes, MME represents one of the most critical entities for the control of the network, since it manages the UEs authentication and authorization. The lack of control over these activities is considered a strong limit for PS networks. As a consequence, we base our proposed architecture on MOCN configuration.

TABLE 1. Types of cells of the proposed HetNet architecture.

Type	Shared-BS	fixed small cells	vehicular cells	cells on wheels or trucks
Sharing	yes	no	no	no
Coverage	macro and micro cells	micro, pico and femto cells	pico and femto cells	micro, pico and femto cells
Availability	always	always	vehicular drive time	vehicular drive time
Power	electrical grid	electrical grid	limited to vehicle battery	generator
Backhaul	wired	wired	wireless	wireless

PS and commercial operators have two different evolved packet cores (EPCs,) and autonomously provide specific services and management policies to their UEs. The two EPCs are connected with a *shared RAN*, and hence with *shared BSs* (S-BSs), by using separate S1 interfaces thus enabling customization. Moreover, the RAN is enriched with small cells dedicated to PS UEs (*dedicated BS*, D-BS) according to the concept of HetNets. This allows to achieve a disaster-resilient and highly-available network architecture. Indeed, PS operator exploits the national-wide RAN deployed by the commercial operator on the basis of suitable service level agreements (SLAs), and increases access points density in some critical areas adding BSs whose access is reserved to PS users, that means PS operators do not have to compete with commercial UEs for BS resources in D-BS. In particular, two different types of D-BSs are foreseen:

- **Fixed small cells** micro, pico and femto cells can be placed by PS operator in strategic/critical locations where PS system requires additional communication capacity and a total control of the network access. For example, this is the case of critical infrastructures or areas where due to the low density of the population the commercial operator is not interested in providing full coverage. Moreover, pico and femto cells can be used in critical areas to improve indoor coverage. These cells are usually characterized by a wired (i.e., high capacity) backhaul links and are placed following suitable planning strategies.
- **Deployable small cells** micro, pico and femto cells that are temporarily deployed to avoid malfunctions and outages in temporally congested area or where it is expected an occasional event (e.g. public event, mass of people) or where terrestrial coverage becomes unavailable due to natural or man-made disasters. Deployable small cells can be vehicles, cells on light trucks and cells on wheels. Cells on vehicles can be immediately available for a prompt response also to very short emergency duration. However, power consumption is limited by the vehicle battery, and hence, coverage is limited to a femto/pico cell. Conversely, cells on trucks or on wheels allow higher coverage because these are usually equipped with power suppliers, but deployment time increases. In all cases, backhaul connectivity should be wireless via dedicated links or via the closest BSs.

The HetNet layers are illustrated in Fig.1, and Table 1 reports the main characteristics of the considered BSs.

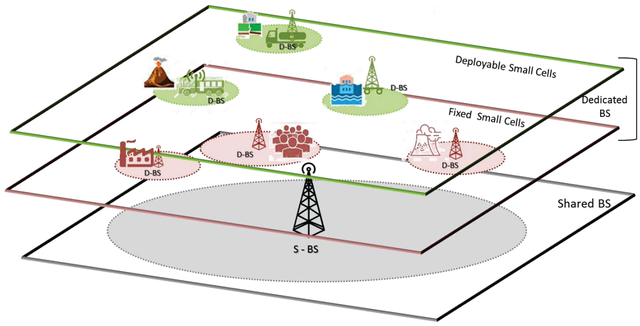


FIGURE 1. HetNet layers.

The proposed architecture is sketched in Fig.2. PS-EPC has two connections with both shared-RAN and dedicated-RAN while commercial EPC is connected only with shared-RAN because D-BSs do not support commercial UEs. The shared-RAN is able to steer the traffic toward the correct EPC using a slice ID.[1] BSs are interconnected through backhaul signaling links using standard LTE X2 interface, that supports a direct control and data information exchange. In particular, control data allows coordination among BSs in order to limit or avoid interference among different network layers, and to manage shared resources. Toward this goal, we assume a centralized RAN management placed in the S-BS that has in charge the coordination among different network layers (i.e., among S-BS and D-BSs in its coverage area) for resources use, and that manages the resources in the shared-RAN. Coordination between different layers is in charge of the Layer Resource Manager (LRM), while to manage the shared RAN we introduce network slicing and resource virtualization that are envisaged as promising and efficient solutions [22]. In particular, we consider two slices: Commercial and PS (*C-Slice* and *PS-Slice*). This means that each S-BS has two *slice resource managers* (SRMs) that schedule their UEs over virtual resources. Then a *resource manager* (RM) maps the virtual resources in physical resources. This architecture recalls basic LTE architecture with the addition of three elements LRM, SRM and PRM that creates a *three-tier resource scheduler*. These elements are described in what follows by introducing examples of scheduling algorithms that could be used in different tiers. However, it is important to stress that the proposed architecture can support different scheduling policies without modifications.

[1] This ID could be encoded in the UE SIM.

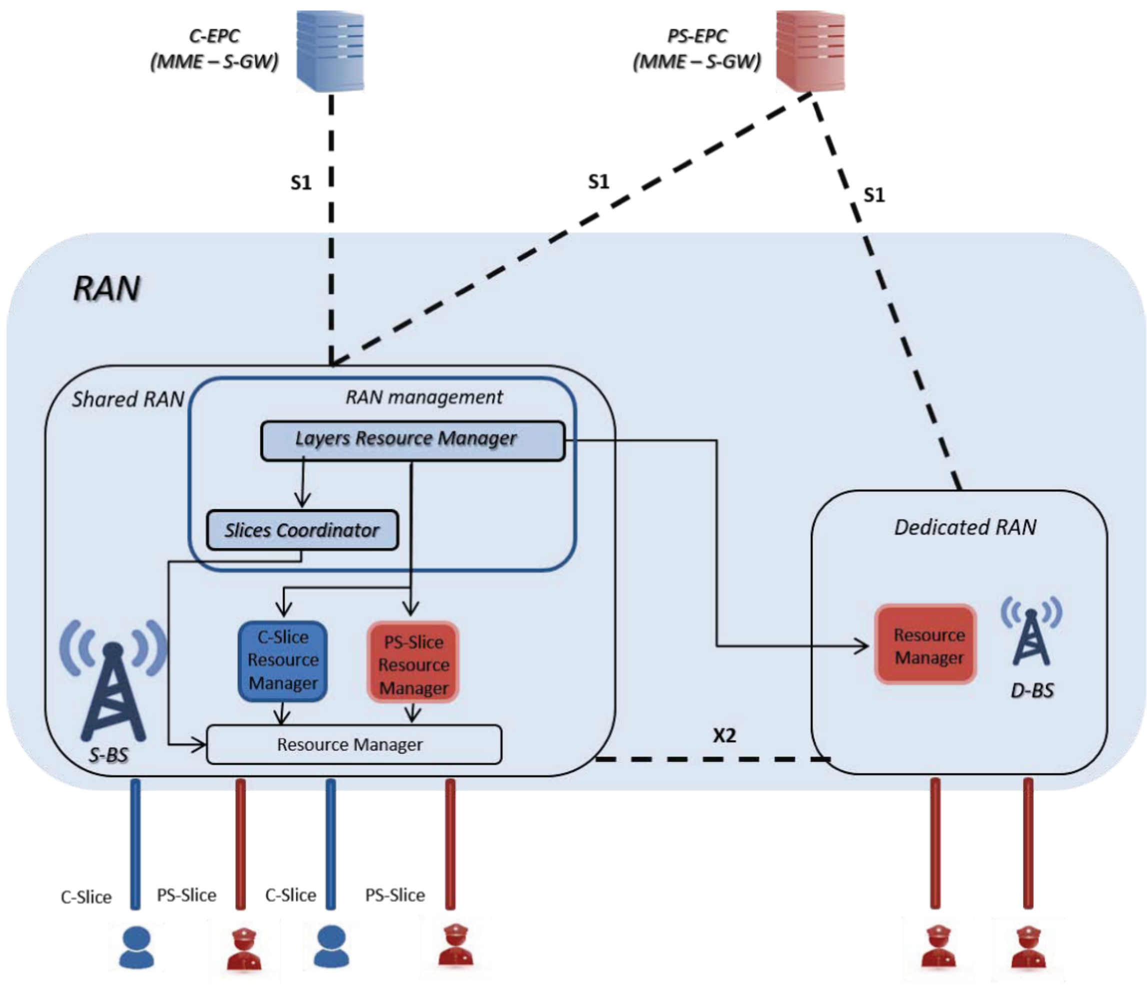


FIGURE 2. **Network architecture.**

C. PROPOSED THREE-TIER RESOURCE SCHEDULER

The proposed architecture pursues the sharing of both RAN infrastructure and radio resources between commercial and PS operators.

In particular, radio resources are managed by using a three-tier scheduler, that exploits different levels of resources' granularity. In particular, to be compliant with LTE standard, we assume that the available resources are organized in frames divided in N subframes, each one composed of n physical resources blocks (PRBs).[2] The scheduler tiers work as follows

- *Tier 1* - is represented by the LRM that assigns radio resources to different HetNet layers (Fig.1) in order to avoid or limit the inter-layer interference while maximizing a target network utility. This tier works at subframe's level. It requires BSs coordination and is periodically performed in the central RAN manager. However, the signaling exchange is limited to the UEs' resource requests and channel quality measurements, moreover, it can be executed only when significant traffic changes occur in the cells. The LRM optimization procedure is performed with a frequency that is set depending on the scenario, and can be also event-driven.
- *Tier 2* - is represented by SRMs that schedule their UEs on virtual resource blocks (VRBs) following their own allocation policies without taking into consideration the other RAN slice; this tier is present only in S-BSs and not in D-BSs.
- *Tier 3* - is represented by RM that maps VRBs coming from SRMs into PRBs. Indeed, the total number of VRBs allocated by SRMs can exceed the number of available PRBs, hence these must be accommodate taking into account the policy indicated by the *slice coordinator*. This policy expresses the percentage of resources to be allocated to each slice, hence the size of the network slices. It changes with the traffic and the context and has to take into account SLAs between operators. In D-BSs RM works as a classical scheduler (i.e., no RAN slicing is present), hence our attention is devoted to S-BS in what follows. In Tier 3 as well

[2] A PRB is the smallest resource unit that can be allocated.

as in Tier 2, resource allocation algorithms run every scheduling interval.

The resource scheduling performed by SRMs and RMs is preferred to a single joint scheduler (only one step without RAN slicing and resource virtualization), because the latter is very complex requiring multi-dimensional scheduling to satisfy different requirements for each slice. Differently, our approach allows more flexibility and scalability, moreover, lets each slice to adopt its own scheduling policy allowing higher customization for the operators.

As stated before, in D-BS the RM works as a classical scheduler allocating its UEs in the subframes indicated by the LRM. However, if during a disaster the shared-RAN is completely destroyed, D-BS can be configured for working in stand-alone mode (for example if the LRM fails in sending inputs for a given period). In stand-alone mode the D-BS uses all the subframes within a frame. Moreover, we assume that PS operator has permanently allocated a limited portion of spectrum to fulfill communications without competition with commercial ones. These resources can be used for providing at least basic narrowband critical services and/or for guaranteeing interference-free communications to the D-BS when coordination among cells is limited or unavailable (i.e., for example deployable small cells with limited backhaul capacity). Obviously, these dedicated resources are not managed by the shared RAN.

D. THREE-TIER SCHEDULING ALGORITHMS

In order to give the proof of concept of our proposed architecture, in this section we introduce some scheduling algorithms suitable for use in each tier. We would like to stress that our goal here is only limited to highligth the effectiveness, efficiency and flexibility of the proposed network architecture, rather than the proposal of the scheduling algorithms. Indeed, the proposed architecture is general and, hence, able to support different scheduling algorithms designed in order to satisfy specific needs, SLAs, types of services and implementation complexity constraints.

1) TIER 1 - LAYER RESOURCE MANAGER

LRM must suitably divide the resources between S-BS and D-BS, hence we consider here an approach to balance the needs of different cells based on an adaptive time division of the resources [8]. The basic idea is to assign to each BS (S-BS and D-BS) the exclusive use of some subframes for communicating with the UEs that would be highly affected by the interference generated by the other BS. In addition, other subframes are used simultaneously by both BSs for communicating with UEs that are in good positions and do not suffer of a high interference.[3] This increases the resource reuse exploiting the concept of enhanced InterCell Interference Coordination (eICIC) [23]. Hence, subframes within a frame are divided in three sets:

[3]We consider that the UEs are already associated to one of the two BSs.

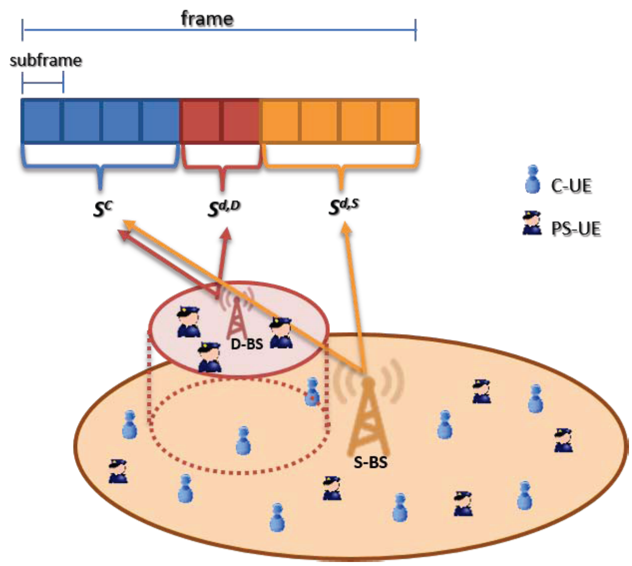


FIGURE 3. Subframes' assignment.

- $\mathcal{S}^{d,D}$ - *dedicated* subframes where only D-BS can transmit data ($|\mathcal{S}^{d,D}| = S^{d,D}$);
- in $\mathcal{S}^{d,S}$ - *dedicated* subframes where only S-BS can transmit data ($|\mathcal{S}^{d,S}| = S^{d,S}$);
- in $\mathcal{S}^{C}$ - *common* subframes where both BSs can communicate simultaneously ($|\mathcal{S}^{C}| = S^{C}$).

As a consequence, we have $N = S^{d,D} + S^{d,S} + S^C$.

Fig. 3 shows the subframes' assignment policy.

Each BS forwards to the LRM the UEs data rate requests, the type of service and the channel quality measurements (with and without interference of the other BS). Based on this information, the LRM is able to calculate the amount of resources requested by each UE both in common and dedicated subframes (i.e., in common subframes the signal-to-noise plus interference ratio -SINR- is lower due to interference, hence, also transmission data rate is lower thus requiring a higher number of PRBs to reach the desired capacity). We indicate with $R_i^S(\mathcal{S}^C)$ and $R_i^S(\mathcal{S}^{d,S})$ the number of PRBs requested by the i-th UE of S-BS in common and dedicated subframes, respectively. In particular, we have

$$R_i^S(\mathcal{S}^C) = \lceil \frac{Q_i}{B\log_2(1 + G_i(\mathcal{S}^C))} \rceil$$
$$R_i^S(\mathcal{S}^{d,S}) = \lceil \frac{Q_i}{B\log_2(1 + G_i(\mathcal{S}^{d,S}))} \rceil \quad (1)$$

where B is the bandwidth of a PRB, Q_i is the capacity amount requested by the i-th UE associated to the S-BS, $G_i(\mathcal{S}^C)$ and $G_i(\mathcal{S}^{d,S})$ represent the averaged SINR in common and dedicated subframes, respectively. The average SINR value takes into account all the link budget parameters, as transmitting power, transmitting and receiving antenna gains and propagation loss. Similar expressions can be defined for the UEs associated to the D-BS, where we have $R_i^D(\mathcal{S}^C)$ and $R_i^D(\mathcal{S}^{d,D})$.

The LRM derives the values of $S^{d,D}$, $S^{d,S}$, and S^{C} in order to minimize the total amount of unsatisfied requests (UR), that is: a UE receives a number of PRBs lower than those needed to reach the requested capacity. In particular, the i-th UE has an amount of UR given by

$$UR_i = Q_i - P_i \tag{2}$$

where P_i is the actual data rate assigned to the i-th UE that depends on the number of PRBs assigned to this UE and its SINR value.

Actually, the URs of the two BSs are suitably weighted to give higher relevance to one of the two BSs depending on the scenario. For example, in case of emergency, D-BS can receive higher priority. Hence, the frame partitioning is derived as

$$\min_{S^C, S^{d,S}, S^{d,D}} (1-\alpha)\sum_{i=1}^{K^S} UR_i + \alpha \sum_{j=1}^{K^D} UR_j \tag{3}$$

where K^D and K^S represent the number of UEs associated to the D-BS and to the S-BS, respectively and $0 \leq \alpha \leq 1$ is the priority weight. This weight represents the sharing policy between the two operators and depends on both the scenario and SLAs. The value of α could be fixed (for example $\alpha = 1$ gives always the highest priority to the PS system) or suitably adapted to the scenario evolution, however how it is managed is out of the scope of the paper.

The proposed algorithm is iterative. The initial state is $S^C = N$ and $S^{d,S} = S^{d,D} = 0$. Then, at each iteration the LRM evaluates the amount of weighted URs (wURs) of the two BSs and:

- if the wURs of S-BS is higher than that of D-BS, $S^{d,S}$ is incremented by one while S^C is decremented by one;
- if the wURs of D-BS is higher than that of S-BS, $S^{d,D}$ is incremented by one while S^C is decremented by one.

Iterations are repeated until $S^C = 0$. This corresponds to N iterations, and to N different subframes configurations. The LRM selects the one that minimizes the total amount of wURs. This means that the LRM algorithm always starts and ends in fixed subframe configurations, while the other $N-2$ configurations are dynamically determined by the algorithm depending on traffic load of the two cells. The output of the algorithm is not the final configuration but the configuration that achieves the lowest amount of wURs.

In order to derive the value of the wURs, UEs are divided in two groups: the first is allocated in dedicated subframes while the second in common subframes. For each UE is calculated the relative increment of the number of PRBs requested if allocated in common subframes instead of in dedicated subframes (i.e., $I = \frac{R_i^{\diamond}(\mathcal{S}^C)-R_i^{\diamond}(\mathcal{S}^{d,\diamond})}{R_i^{\diamond}(\mathcal{S}^C)}$ where $\diamond = S, C$ depending the cells to which the UE belongs to), and then UEs are listed in descending order. The first $X^{\diamond}$ UEs belonging to $\diamond$-BS are allocated in dedicated subframes ($\mathcal{S}^{S(d,\diamond)}$), while the remaining are allocated in common subframes ($\mathcal{S}^C$). The value $X^{\diamond}$ is calculated so that the amount of total URs is almost the same in both portions of the frame.

We want to underline that LRM does not distinguish between commercial and PS UEs, but only between UEs associated with D-BS and S-BS.

The output of the LRM provided to the other scheduler tiers (in particular to the slice coordinator and SRMs) is represented by the frame repartition ($S^{d,S}$, S^C, $S^{d,D}$), and the indication if the UE must be allocated in a common or dedicated subframe.[4]

The proposed LMR algorithm is presented for a scenario with one D-BS and one S-BS. However, in the case of more than one S-BS in the coverage area of the macrocell the algorithm still works. The subframes are always divided in the three sets: a common subframe division for multiple small cells is used. Therefore, all small cells simultaneously communicate on the same $S^{d,D}$ subframes.[5] What changes is the LMR input, indeed the subframes division should be evaluated for example by taking into account the small cell with the highest traffic load or a traffic load averaged on all the small cells.

2) TIER 2 - SLICE RESOURCE MANAGER

Within S-BS, UEs belong to two slices (Commercial and PS) and are independently allocated in the virtual resources by the two SRMs. In particular, in each SRM the number of available VRBs is equal to the total amount of RBs available at the S-BS ($\mathcal{S}^{d,S} \cdot n + \mathcal{S}^C \cdot n$). That is, each SRM may schedule the UEs belonging to its class running up all the S-BS resources. However, when the VRBs are mapped in PRBs (in RM), physical limits will be taken into account. SRM can use any scheduling algorithm (such as Proportional Fair, Round-Robin, Priority-Based, Delay-Based), and each SRM may use a different scheduling policy. We have considered here a Proportional Fair (PF) approach for both SRMs, but this is only an example to show the effectiveness of the proposed architecture. In particular, PS-SRM first allocates all UEs that request a high-priority (critical) service, while the remaining UEs are allocated following the PF approach. Each UE receives an amount of VRBs that is proportional to its requested capacity. Similarly, C-SRM schedules its UEs with a PF approach. We assume here two different classes of service for PS UEs and only one for commercial UEs. Both SRMs communicate to the RM the amount of VRBs allocated to each UE with some *attributes*. In our example the attributes are the priority of the UE and the type of subframe (dedicated or common).

3) TIER 3 - RESOURCE MANAGER

RM receives as inputs the virtual resource allocations performed by the two SRMs and the sharing policy indicated by the slice coordinator. We assume that the slicing coordinator provides to the RM a weight (Γ) that is used to balance the PRBs assignment. Hnce, $\Gamma \in [0, 2]$ is a system parameter

[4] With the relative modulation and coding scheme.

[5] The low power and coverage of the small cells limits the intra-layer interference.

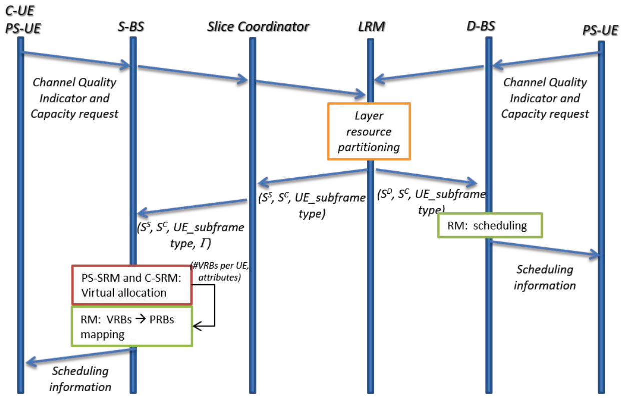


FIGURE 4. Messages exchange and functions of three-tier scheduler.

that takes into account SLAs between the two operators and also the context. For example, in case of emergency the access priority of the PS UEs must be increased. The RM algorithm runs independently on how the value of Γ is derived.

The RM first maps all VRBs requested by high-priority UEs in PRBs belonging to the subframes indicated by the upper tiers, then maps the VRBs of the remaining UEs using a proportional approach (i.e., the number of PRBs are calculated proportionally to Γ). In particular, if $\Gamma = 2$ all the VRBs of PS-UEs are first mapped in PRBs and the remaining are used for commercial UEs. Viceversa if $\Gamma = 0$.

The three-tier scheduling messages exchange is pictured in Fig. 4.

V. NUMERICAL RESULTS

In this section the performance of the three-tier scheduler is analysed in order to give a proof of concept of the proposed architecture.

Numerical results have been derived by means MATLAB simulations, in a scenario with one S-BS and one D-BS, whose equivalent isotropic radiated power (EIRP) is $41dB$ and $20dB$, respectively. UEs and D-BS are randomly distributed in the D-BS coverage area. To make final results independent from one particular distribution of the UEs and D-BS, multiple trials have been performed and averaged.

We consider three different scenarios in accordance with [24]: *(i) normal (N)*, *(ii) peak (P)*, *(iii) emergency (E)*. The *normal* case refers to routine work, while the *peak traffic* case can temporary happen with an increment of about 30% of average traffic. *Emergency* case occurs when one or more critical events happen, new services are activated, while others are less used, but in general both the number of PS operators and needed resources per user grow. Tab.2 reports the traffic characteristics considered in the simulations derived from [24]. Moreover, we have considered a variable number of commercial users (C-UEs) in the range [50, 200], a total bandwidth of 10 MHz, and a number of subframes per frame $N = 10$. The total frame duration is assumed equal to 10ms. For what concerns the path-loss the Hata-Cost 231 model is considered.

TABLE 2. PS data traffic: critical, real-time, no real-time (NRT).

Scenario	Service	Number UEs	Data rate	Priority
Normal	Critical voice	30	16kbps	1
	Real-Time	10	300kbps	2
	NRT Narrowband	10	80kbps	2
	NRT Broadband	5	800kbps	2
Peak	Critical voice	55	16kbps	1
	RT broadband	10	300kbps	2
	NRT Narrowband	15	80kbps	2
	NRT Broadband	5	800kbps	2
Emergency	Critical voice	70	16kbps	1
	RT broadband	20	300kbps	2
	NRT Narrowband	30	80kbps	2
	NRT Broadband	15	800kbps	2

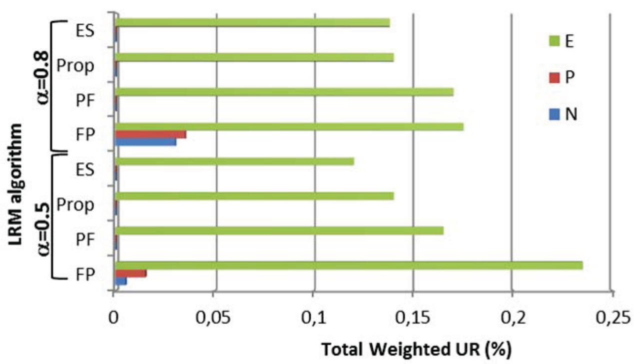
First of all, we show the performance of LRM. Fig.5 represents the total wURs normalized to the total resource requests for multiple scenarios and values of α varying the number of C-UEs whose traffic is classified as:

- 65% of the UEs 16 kbps;
- 25% of the UEs 160 kbps;
- 10% of the UEs 500 kbps.

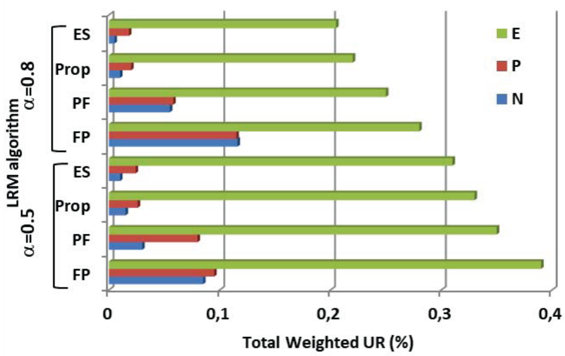
In order to show the effectiveness of the proposed scheme (indicated as *Prop* in the figures) we introduce three different benchmark approaches. The optimal solution of (3) achieved by an *Exhaustive Search* (ES) is reported in order to show the accuracy of the proposed heuristic. Moreover, we have considered the *fixed partitioning* (FP) method that foresees an orthogonal and fixed resource partitioning between D-BS and S-BS without shared subframes (i.e., $S^C = 0$). The number of subframes dedicated to S-BS and D-BS is calculated proportionally to the α value as: $S^{d,D} = round(\alpha N)$ and $S^{d,S} = N - S^{d,D}$. This comparison method allows to show the flexibility and the advantages of resource reuse of our method. Finally, to show the effectiveness of the proposed HetNet architecture, we consider a single BS providing service in the overall area. UEs are scheduled following a *PF* approach that takes into account both the UE's capacity request and its priority (i.e. α).

From Fig. 5 it is evident that the resource sharing flexibility, introduced by the proposed LRM, allows a significant improvement of the performance in comparison with FP that is subject to waste of resources (i.e., unused resources by one of the two system while the other has URs). Moreover, the introduction of D-BS guarantees a higher capacity in comparison with a single BS thanks to resources reuse as it is evident from the comparison with the PF method. If the traffic load is not particularly high the proposed LRM as well as PF method obtain almost zero URs value, because there are sufficient resources to accommodate all the UEs' requests. But when the number of commercial UEs increases or in emergency situations the benefits of our method are more evident.

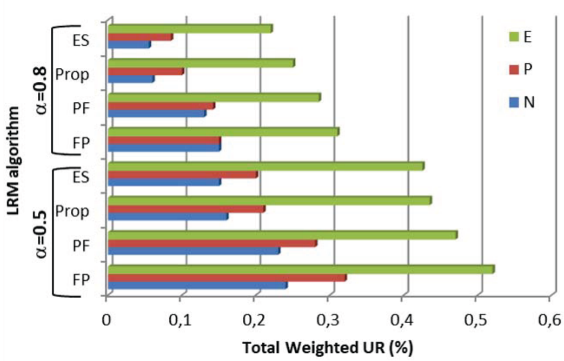
The results also show that the normalized URs value of the proposed method is very close to that of the ES, thus validating the accuracy of the heuristic that allows a significant reduction of the complexity. Indeed, ES needs to calculate the wURs for $\sum_{i=1}^{N+1} \frac{(N+1)(N+2)}{2}$ configurations, in comparison with N of our method.



(a)



(b)



(c)

FIGURE 5. Total amount of weighted URs. (a) 50 C-UE. (b) 100 C-UE. (c) 200 C-UE.

In order to highlight the effect of weight α, in Fig. 6 the URs of each BS normalized to the relative total data rate request, are showed for a *peak* scenario considering two values of α when the number of C-UEs in the area changes. It is evident that under the same traffic load condition, changing the value of the weight allows to adapt the system to the specific needs. Indeed, when a PS-related event occurs it can be useful to increase the value of α so that the percentage of URs for the PS systems remains very close to zero even in presence of a high load. Obviously, this results in a worsening of the commercial network performance.

Finally, we want to show the effectiveness of RAN slicing and resource virtualization by means a comparison with a

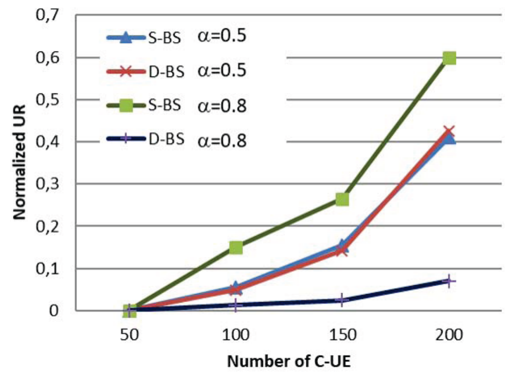


FIGURE 6. Normalized URs of the two BSs for different values of α.

TABLE 3. URs using RAN slicing and virtualization (VIR) or physical allocation (PHY).

	Total wURs		PS-URs		C-URs	
Γ	1	1.8	1	1.8	1	1.8
PHY	5.5%	0.7%	0%	0%	8.5%	9.4%
VIR	4%	0.55%	4%	0%	4%	6.8%

physical resource allocation (without slicing) that divides the available PRBs in two sets proportionally to Γ, one for PS-UEs and the other for C-UEs. In this way, similarly to our approach, allocation is separately managed for the two types of UEs in each portion of resources. However, working on PRBs instead of VRBs leads inefficiencies and waste of resources. In Tab. 3 we report the results in terms of normalized URs achieved for a *peak* scenario, assuming almost the same number of C-UEs and PS-UEs in the area, for two different values of Γ. We can see that using RAN slicing and resources virtualization the URs decreases because the system is more flexible. If it is mandatory to assure no-URs to the PS system, the value of Γ is set close to 2, and using virtual allocation it is possible to reduce also the commercial URs amount with a more efficient resource exploitation. Conversely, setting an equal priority for the two networks (i.e., $\Gamma = 1$), with virtual allocation there is a fairer distribution of URs among the two networks and an overall reduction of the URs.

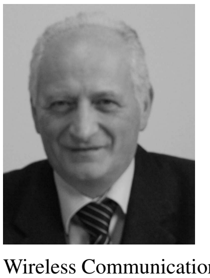
VI. CONCLUSION

This paper proposed a new architecture based on Heterogeneous Network and Spectrum Sharing paradigms. Dedicated small cells are deployed where and when needed to complete the basic commercial shared network in order to satisfy PS stringent requirements in terms of coverage, capacity and resilience. This architecture is managed by means a three-tiers scheduler that (i) allows a suitable resource partitioning between dedicated and shared base stations, (ii) introduces RAN-sharing and network virtualization to efficiently manage resource allocation in the shared base stations, while maintaining independent allocation policies for commercial and public safety users without resorting to very complex multi-dimensional schedulers.

REFERENCES

[1] T. Doumi *et al.*, "LTE for public safety networks," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 106–112, Feb. 2013.
[2] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahmi, "Device-to-device communications for national security and public safety," *IEEE Access*, vol. 2, pp. 1510–1520, Dec. 2014.
[3] L. Carlà, R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, "LTE enhancements for public safety and security communications to support group multimedia communications," *IEEE Netw.*, vol. 30, no. 1, pp. 80–85, Jan. 2016.
[4] A. Kumbhar, F. Koohifar, I. Güvenç, and B. Mueller, "A survey on legacy and emerging technologies for public safety communications," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 97–124, 1st Quart., 2017.
[5] M. Usman, A. A. Gebremariam, U. Raza, and F. Granelli, "A software-defined device-to-device communication architecture for public safety applications in 5G networks," *IEEE Access*, vol. 3, pp. 1649–1654, 2015.
[6] K. Ali, H. X. Nguyen, P. Shah, Q.-T. Vien, and N. Bhuvanasundaram, "Architecture for public safety network using d2d communication," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.
[7] A. Al-Hourani and S. Kandeepan, "Temporary cognitive femtocell network for public safety LTE," in *Proc. IEEE 18th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Sep. 2013, pp. 190–195.
[8] G. Bartoli, R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, "Dynamic PSS and commercial networks spectrum sharing in the time domain," *Trans. Emerg. Telecomm. Technol.*, vol. 28, no. 3, p. e3083, Mar. 2017.
[9] R. Favraud, A. Apostolaras, N. Nikaein, and T. Korakis, "Toward moving public safety networks," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 14–20, Mar. 2016.
[10] M. Shin, S. T. Shah, M. Y. Chung, S. F. Hasan, B.-C. Seet, and P. H. J. Chong, "Moving small cells in public safety networks," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2017, pp. 564–568.
[11] X. Chen, D. Guo, and J. Grosspietsch, "The public safety broadband network: A novel architecture with mobile base stations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 3328–3332.
[12] A. Merwaday, A. Tuncer, A. Kumbhar, and I. Guvenc, "Improved throughput coverage in natural disasters: Unmanned aerial base stations for public-safety communications," *IEEE Veh. Technol. Mag.*, vol. 11, no. 4, pp. 53–60, Dec. 2016.
[13] M. M. Sohul, M. Yao, X. Ma, E. Y. Imana, V. Marojevic, and J. H. Reed, "Next generation public safety networks: A spectrum sharing approach," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 30–36, Mar. 2016.
[14] M. Yuksel, I. Guvenc, W. Saad, and N. Kapucu, "Pervasive spectrum sharing for public safety communications," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 22–29, Mar. 2016.
[15] H. Shajaiah, A. Abdel-Hadi, and C. Clancy, "Spectrum sharing between public safety and commercial users in 4G-LTE," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 674–679.
[16] Q. Wang and T. X. Brown, "Public safety and commercial spectrum sharing via network pricing and admission control," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 622–632, Apr. 2007.
[17] S. Borkar, D. Roberson, and K. Zdunek, "Priority access for public safety on shared commercial LTE networks," in *Proc. Tech. Symp. ITU Telecom World (ITU WT)*, Oct. 2011, pp. 105–110.
[18] R. Ferrus, O. Sallent, G. Baldini, and L. Goratti, "Public safety communications: Enhancement through cognitive radio and spectrum sharing principles," *IEEE Veh. Technol. Mag.*, vol. 7, no. 2, pp. 54–61, Jun. 2012.
[19] A. Somov, T. Rasheed, and V. K. Yedugundla, "Power control game for spectrum sharing in public safety communications," in *Proc. IEEE 18th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Sep. 2013, pp. 207–211.
[20] R. Fantacci, F. Gei, D. Marabissi, and L. Micciullo, "Public safety networks evolution toward broadband: Sharing infrastructures and spectrum with commercial systems," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 24–30, Apr. 2016.
[21] *Technical Specification Group Services and System Aspects; Network Sharing; Architecture and Functional Description (Release 10)*, document 3GPP TS 23.251 V10.6.0, Jun. 2013.
[22] G. Tseliou, F. Adelantado, and C. Verikoukis, "Scalable RAN virtualization in multitenant LTE—A heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6651–6664, Aug. 2016.
[23] G. Bartoli, R. Fantacci, D. Marabissi, and M. Pucci, "Adaptive muting ratio in enhanced inter-cell interference coordination for LTE—A systems," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2014, pp. 990–995.
[24] CEPT-ECC Electron. Commun. Committee, "User requirements and spectrum needs for future European broadband PPDR systems (wide area networks)," Tech. Rep. ECC 199, 2013.

DANIA MARABISSI (SM'13) received the master's degree *summa cum laude* in telecommunication engineering in 2000 and the Ph.D. degree in informatics and telecommunications engineering from the University of Florence in 2004. She is currently an Assistant Professor with the University of Florence. She has authored several technical papers published in international journals and conferences and patents. Her research activity concerns mainly with new generation wireless communication systems with particular attention to physical and access layers issues. This activity has been carried out in national and international funded projects and in collaboration with companies of the sector. She is the winner of the contest FIR-Futuro in Ricerca 2013 funded by the Italian Ministry of Education, with the Project Heterogeneous LTE Deployment. She is also responsible of the scientific activities of the Technologies for Information and Communication consortium, born between the University of Florence and Leonardo Company S.p.A., with the aim to promote collaboration and technological transfer among university and industry. She is an Associate Editor of the IEEE Transaction on Communications, the IEEE Transactions on Vehicular Technology, the *Transaction on Emerging Telecommunication Technologies*, the IET Communications and *The Scientific World Journal* (Hindawi) and has been involved in the organization of several international conferences.



ROMANO FANTACCI (F'05) received the M.S. degree in electrical engineering from the University of Florence, Italy, and the Ph.D. degree in computer networks from the University of Florence, Italy. He is currently a Full Professor of computer networks with the University of Florence, Florence, Italy, where he heads the Wireless Networks Research Lab. He is the Founding Director of the Information Communication Technology Inc., with Finmeccanica spa and of the Wireless Communications Research Center with Telecom Italia. His current research interests encompass several fields of wireless engineering and computer communication networking including, in particular, performance evaluation and optimization of wireless networks, emerging generations of wireless standards, cognitive wireless communications and networks, satellite communications and systems. He was an elected fellow of the IEEE in 2005 for contributions to wireless communication networks. He received several awards for his research, including the IEE Benefactor Premium, the 2002 IEEE Distinguished Contributions to Satellite Communications Award, the 2015 IEEE WTC Recognition Award, the IEEE sister society AEIT Young Research Award and the IARIA Best Paper Award, the IEEE IWCMC'16 Best Paper Award, and the IEEE Globecom'16 Best Paper Award. He served as an Area Editor for the IEEE Transactions Wireless Communicatons, an Associate Editor for the IEEE Transactions on Communication, the IEEE Transactions Wireless Communication, and an Associate Editor for the *Telecommunications Systems*. He guest edited special issues for the IEEE journals and magazines and served as the symposium chair of several IEEE conferences, including VTC, WCNC, PIRMC, ICC, and Globecom. He currently serves on the Board of Governors of the IEEE Sister Society AEIT, as an Associate Editor for the *International Jounal Communication Systems*, and as a Regional Editor for the IET Communications.