SoftPSN: Software-Defined Resource Slicing for Low-Latency Reliable Public Safety Networks

Anteneh A. Gebremariam, Muhammad Usman, Riccardo Bassoli, and Fabrizio Granelli

Department of Information Engineering & Computer Science (DISI), University of Trento, Trento, Italy

Correspondence should be addressed to Fabrizio Granelli; fabrizio.granelli@unitn.it

Received 30 December 2017; Accepted 11 April 2018; Published 14 May 2018

1. Introduction

The vision of future 5G cellular networks and beyond (B5G) has added various use cases to the current 4G cellular networks. Instead of sole mobile broadband, the aim is to provide full support for new technologies such as augmented reality, Internet of Things (IoT), Internet of Vehicles (IoV), Tactile Internet, and Machine-to-Machine (M2M) and Machine-to-Human (M2H) communications [1]. In parallel, the future 5G ecosystem [2] will also support governmental and defence communications, communications for disaster relief, humanitarian aid, and intelligent transportation systems. That means the plan is to make 5G the main infrastructure for future generations of public safety networks (PSNs).

5G and B5G will need to address several challenges to guarantee the desired key performance indicators (KPIs). In particular, these requirements are as follows:

(i) Throughput: 10 times higher data rates to individual end users.

(ii) Latency: 1–10 ms round-trip time to allow also the support for URLLCs.

(iii) Higher bandwidth per unit area and enormous number of connected devices.

(iv) Perceived network availability of 99.999% to allow also the support for URLLCs.

(v) Energy efficiency: energy-usage reduction of about 90%.

(vi) Service creation time: time to set up a service from application to individual service components in a second or less.

(vii) Battery lifetime: 10 times increase in battery life.

The scope of public safety networks is the provisioning of effective communications infrastructure in case of emergency/catastrophic events, which either can significantly reduce network coverage/capabilities or can create a complete outage. During these events, the first responders such as law enforcement, firefighters, emergency medical staff, the military, and volunteer organizations should have a robust and efficient network to rely on. In particular, Professional Mobile Radio (PMR) and Public Safety Land Mobile Radio (PLMR) are the current technologies used by public safety entities. Moreover, current PSNs can be categorized into...
two main families: LMSR (i.e., TETRA and APCO-25) and broadband networks (i.e., LTE for PSN such as FirstNet).

In the last decade, significant part of the research community focused on enhancing PSNs [5, 6]. The investigation has released several solutions mainly involving massive MIMO, small cells, unmanned aerial vehicles (UAVs), satellite networks, ad hoc networks, spectrum sharing, and cognitive radio.

Even if there is a wide panorama of solutions to enhance PSNs, actual assessments of disaster response capabilities in terms of response time and available capacity have underlined the limitations [7] of current PSNs and the open issues in case of real disasters (e.g., Japan Earthquake 2011). The main problems of that inefficiency are due to the following:

(i) Lack of interoperability among networks of different rescue/emergency entities/organizations; no interoperability with information provided by normal users during and after the disaster.

(ii) High error rate due to outage and congestions; keep network robust.

(iii) Inefficiency to provide significant communication resources after the disaster; traffic congestion due to rapid traffic generation. Ten times larger traffic with low percentage of resources available. The round-trip time (RTT) after disaster could reach 100–150 ms and even hours in the worst cases.

(iv) No possibility of providing television/radio broadcasting support after disaster.

Thus, future 5G networks can provide the right infrastructure to address all the above problems via dedicated interfaces for ultrareliable low-latency communications (URLLCs), supported by efficient virtualization, slicing, and dynamic adaptation: this can guarantee an effective amount of bandwidth and significantly lower delays under critical conditions as well.

During the last years, part of the research tried to improve current PSNs by targeting the lack of coverage and bandwidth with the deployment of UAVs [8]. In particular, the use of unmanned aerial base stations (UABSSs) can have the advantage of rapid deployment and large on-demand coverage provisioning. However, battery lifetime and interference management become crucial issues. Moreover, after a disaster, the correct setup of UABS-based access network still requires some time and does not address the immediate need of optimized resources.

This article provides an efficient and effective virtualized architecture to address URLLCs provisioning to PSNs in the presence of scarce resources: that may be due to a natural disaster, a terrorist attack, or other catastrophic events. In particular, the solution employs virtualization of resources at eNodeB, performing slicing operations, thus enabling first responders to use the existing scarce radio resources in an effective way until more resources are available via the deployment of UABSSs. Moreover, short-range and device-to-device communications provide efficient additional coverage and bandwidth while guaranteeing low delays. The whole architecture combines all the aforementioned technologies to guarantee robust communications and RTTs in line with 5G KPIs for URLLCs.

The main contributions of this paper include the following:

(i) A novel architecture based on virtualization of network resources; a resource slicing algorithm tailored to handle the radio resource insufficiency to guarantee optimal integration of PSNs into current LTE cellular networks.

(ii) A novel solution to integrate software-defined networking (SDN) and short-range communications among user equipment to guarantee PSNs’ operational requirements in the context of 5G-based PSNs.

(iii) Simulations and numerical evaluations of the overall system to prove that URLLCs in PSNs can achieve the desired latency.

Moreover, the article introduces the basics to apply stochastic geometry to model the behavior of cellular networks after a disaster.

The paper is structured as follows. Section 2 introduces the system model and describes in detail the network structure and the architecture. It also provides a clear description of how slicing virtual resources works. Next, Section 3 discusses how optimization of resources works and how it can provide optimal coverage/capacity when scarce. Section 4 describes how the slicing of resources and the D2D communications are working together for optimal resources provisioning, towards low-latency communications. Finally, Section 5 shows and discusses the achieved results in terms of latency and efficiency. Furthermore, it gives some direction to guarantee the security of the D2D environment.

2. System Model

The following subsections introduce in detail the assumptions and hypotheses under which the system can provide URLLCs to public safety networks. Side by side, they discuss the network and the radio resource slicing models.

2.1. Network Model

Let us consider an urban scenario with a medium/high device density (greater than $10^3$ devices per km$^2$ [9]). Next, at a specific time $t_1$, a general catastrophic event happens so that the infrastructure becomes limited [9]: in particular, that means only a small percentage of macro cell coverage and cellular resources remain available.

In order to provide an accurate model for the scenario, the article borrows the distribution of base stations (BSs) in the dense urban environment of London [10]. Then, the distribution of BSs can be modeled via a Poisson point process (PPP) of density (for time $t < t_1$) $\lambda = N/A = 80[10]$.

Nevertheless, at $t = t_1$, the catastrophic event happens, and the PPP of the BSs changes. This change can be modeled as an independent thinning transformation [11]. In particular, this means removing certain points according to a probabilistic rule, independent of all points.
Let $g : \mathbb{R} \to [0,1]$ be the thinning function applied to the PPP of BSs $\Phi_{BS}$, by deleting each point $x \in \Phi_{BS}$ with probability $1 - g(x)$, independently of all other points. At $t > t_1$, $\Phi_{BS}'$ denotes the inhomogeneous PPP with new intensity $\lambda_{BS} g(x)$.

Figure 1 depicts an example of how a London-like urban scenario changes because of a disaster. Figure 1(a) shows the map of the coverage, given the full network. Figure 1(b) represents the Voronoi tessellation generated by the thinned PPP. The disaster is simply modeled by a thinning function that randomly and uniformly selects a percentage of BSs, which remains active.

Next, the probability mass function (PMF) of the amount of user equipment (UE) in a Voronoi cell can be approximated as

$$
\Pr \{ K = k \} = \frac{343}{15} \sqrt{\frac{7}{2\pi k!}} \left( \frac{\lambda_{MT}}{\lambda_{BS}} \right)^k \left( \frac{\lambda_{MT}}{\lambda_{BS}} + \frac{7}{2} \right)^{-k-7/2} \cdot \Gamma \left( k + \frac{7}{2} \right). \tag{1}
$$

Let us assume a channel model, which has unbounded path loss and Rayleigh fading. The power path loss of a BS-to-UE link can be expressed as

$$
\bar{t}(r) = \frac{4\pi}{v} r^{-\alpha}, \tag{2}
$$

where $r$ is the distance of MT from BS, $4\pi v$ denotes the free space path loss (FSPL) at 1 meter from the transmitting device ($v$ is the wavelength), and $\alpha > 2$ is the path loss exponent.

Furthermore, a generic BS-to-UE link only experiences flat Rayleigh fading $h$ with expected value $E[h] = \Omega$ and probability density function (PDF):

$$
f_h(x) = \frac{1}{\Omega} \exp \left( -\frac{x}{\Omega} \right). \tag{3}
$$

The fading of each link is independent and identically distributed (i.i.d.), and it is assumed to be constant during a transmission time slot and to independently change for each time slot.

Let BS$_0$ be a general base station in the Voronoi cell V(0), located at the position $x_0$; this Voronoi cell contains $K$ mobile terminals. Let $\Phi_{BS}'$ be the set of base stations $\Phi_{BS} \setminus 0$, which considers all the BSs not in V(0). So, the signal-to-interference-plus-noise ratio (SINR) at UE$_k$ in the Voronoi cell V(0) in time slot $\tau$ becomes

$$
\text{SINR}_k^\tau = \frac{P h_{0,k}(\bar{t}(r_0))^{-1}}{N_0 + P \sum_{i \in \Phi_{BS}'} h_{i,k}(\bar{t}(r_i))^{-1}}, \tag{4}
$$

where $P$ is the transmission power, $N_0$ is the variance of the additive white Gaussian noise (AWGN), $r_0 = \|x_k - y_0\|$, and $r_i = \|x_k - y_i\|$.

Finally, the transmission success probability of a packet during time slot $\tau$ is

$$
p_s = \Pr \{ \text{SINR}_k^\tau > \beta \}, \tag{5}
$$

where $\beta$ is the SINR threshold. In particular, this threshold is related to the transmission rate $R$ at the physical layer, and it can be expressed as $\beta = 2^R - 1$ due to Shannon’s capacity formula.

Hence, the above expressions analytically show that the reduction of BSs implies a new scenario with larger cells, containing more UE to be served, which can be at a significant distance from the serving BS. Furthermore, by keeping the same transmission power, it is possible to see that the SINR decreases together with the quality and the latency of the transmission.

2.2. Resource Slicing Model. As a result of the disaster, base stations are forced to extend their coverage area to accommodate more users as depicted in Figure 1(b), and thus the amount of available radio resources to serve users is reduced. To alleviate this issue, we propose an efficient way of allocating resources to enable reliable connection after
disasters. In this section, we discuss the effects of dynamic radio resource slicing on achieving a low-latency constraint for a public safety scenario.

We consider a system model which is based on the work in [3], Figure 2, where the Slice Manager (SM) sends resource slicing trigger messages to the Radio Resource Slicer (RRS) (i.e., dynamic spectrum-level slicer (DSLS) and Priority-Based Resource Slicer (PBRS)) based on the radio resources available in virtual radio resource pool (VRRP) corresponding to each of the virtual networks (VNs). VRRP is the abstraction of radio resources, a three-dimensional (3D) pool of resources in time, frequency, and space, from the operational base stations as shown in Figure 1. The \( \langle \text{time}, \text{frequency} \rangle \) dimensions represent the orthogonal frequency-division multiplexing (OFDM) resource blocks (RBs) considering LTE, while the \( \langle \text{space} \rangle \) dimension represents the serving base station. The smallest 3D virtual radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot is defined as a virtual RB (vRB). Each VN supports its own service and also allocates radio resource block per slot.

For our analysis, we considered two types of users: (i) primary users (PUs), that is, the first responders belonging to virtual network 1 (VN\(_1\)), and (ii) secondary users (SUs), that is, the rest of the users accessing the network belonging to VN\(_2\). In order to achieve a low latency for URLLCs after a disaster, the PUs need to be granted more radio resources as compared to the SUs. This enables the first responders to transmit/receive the required content in case of scarce radio resources to deliver the required disaster relief. Relying on the network and system models described above, the full details of the radio resource slicing and allocation are presented in the following section.

3. Optimization and Radio Resource Slicing

The main goal of the radio resource slicing and optimization is to give higher priority in case of disaster, in terms of resource allocation, to the first responders in order to fit the low-latency requirement. Otherwise, the resource slicing is just performed as it is proposed in [3]. To achieve this, we modify the resource slicing optimization problem defined in [3] by introducing a resource allocation priority parameter \( \zeta \). In Figure 2, the parameter \( L_{kj} \) represents the traffic load, the total number of radio resources required by all users per available radio resource, in slice (VN) \( j \in \{1,2\} \) in transmission time interval (TTI) \( k \in \mathbb{N} \). When \( L_{kj} > 1 \), the SM sends a trigger message to the Radio Resource Slicer (RRS) to slice the available radio resources into two segments \( (\psi_{k1}) \), which is the optimal radio resource allocation of each slice.

The resource slicing algorithm that is tailored for the public safety scenario is presented by a pseudocode as shown in Algorithm 1. As it can be seen in the algorithm, the priority-based resource slicing is performed only if the traffic load \( L_{kj} \) is too high indicating that some of the available base stations are out of service. On the other hand, if the traffic load is reasonable, the SM decides to use the DSLS algorithm to slice the available radio resources. The initialization, \( \text{Line (4)} \), represents an equal number of vRB allocations for each VN. In \( \text{Line (11)} \), we have the resource schedulers, MAC Scheduler \( j \), that are optimized to the services supported by each of the VNs.

Based on Algorithm 1, the SM (see Figure 2) decides which resource slicing algorithm to use. We describe the details of each resource slicing module as follows.

**Priority-Based Resource Slicer (\( \zeta \)).** Depending on the intensity of the traffic load, we could assume a range of values for \( \zeta \) that could be updated dynamically, that is, \( \zeta \in [0.7,0.9] \), indicating more vRB allocation to VN\(_1\). Thus, the resource allocation for each of the VNs becomes

\[
\psi_{k1} \leftarrow \zeta N_{RB}, \\
\psi_{k2} \leftarrow (1 - \zeta) N_{RB},
\]

where \( \psi_{k1} \) and \( \psi_{k2} \) are the sliced radio resources for VN\(_1\) and VN\(_2\), respectively, in TTI \( k \).
procedure ResourceSlicing
for each simulation transmission time interval \( k \) do
    for each \( j \in \{1, 2\} \) do
        Initialization \( \psi_{0j} \leftarrow N_{RB}/2 \)
    end for
    if \( L_{kj} \gg 1 \) then
        \( \psi_{kj} \leftarrow \text{Priority-based Resource Slicing (\( \zeta \))} \)
    else
        \( \psi_{kj} \leftarrow \text{Perform DSLS [3]} \)
    end if
end for
VN\(_j\) ResourceScheduling (MACScheduler\(_j\))
end procedure

Algorithm 1: Resource slicing algorithm for public safety application.

Table 1: Supported ProSe functions in 3GPP release 12 to enable D2D communication in public safety and non-public safety applications [4].

| Scenarios         | Within network coverage | Outside network coverage | Partial network coverage |
|-------------------|-------------------------|--------------------------|--------------------------|
| Supported applications | Discovery               | Communication            | Communication            |
| Non-public safety | Discovery               | -                        | -                        |
| Public safety     | Discovery, communication| Communication            | Communication            |

DSLS. The optimization problem for the DSLS [3] is targeted to maximize the overall system utility (logarithmic function) subject to a set of constraints as shown below:

\[
\begin{align*}
\text{maximize} & \quad \sum_j \ln \left( \psi_{kj} \right), \\
\text{subject to} & \quad \sum_j \psi_{kj} = N_{RB}, \quad (7) \\
& \quad \psi_{est,j} \leq \psi_{kj}, \\
& \quad \psi_{kj} \geq 0,
\end{align*}
\]

where \( \psi_{est,j} \) represents the total number of vRBs needed to satisfy the current traffic in VN \( j \) in TTI \( k \).

Applying Algorithm 1, we can be able to slice a pool of radio resources that are abstracted in a central entity with a reasonably low computational complexity. In addition, since the resource slicing algorithm will not be executed every TTI, as indicated in [3], we can still be able to perform the computation without exceeding the latency constraint given by each VN’s services.

4. An Example Scenario: A D2D-Based Application

This section discusses the detail of D2D communication in VN\(_j\) among first responders in the emergency scenarios or traffic hotspot situations. MAC Scheduler \(_j\) residing in VN\(_j\) allocates radio resource blocks to first responders to achieve the URLLC latency constraint. To measure the latency of group communication among first responders, we consider a content distribution scenario, where first responders want to share emergency content among each other.

The group communication among first responders has been standardized in 3rd-generation partnership project (3GPP) release 12 under the name of proximity services (ProSe). The scenarios supported by ProSe in 3GPP release 12 are shown in Table 1. In this work, we consider an in-band, in-network coverage, D2D scenario wherein public safety UE is given a high priority over commercial cellular UE. It is important to note that as D2D radio resource allocation and communication are controlled by a cellular network, we do not consider the device discovery phase in estimating the network latency.

In our simulations, we measure the network latency by assuming that resource allocation is performed every TTL (1 ms), while the optimization algorithm is run only when needed. Once a group is formed, the first responders start communicating with each other on D2D links using the allocated resources. A first responder shares an emergency content with the others in the group.

A visual illustration of the considered scenario is shown in Figure 3, where one of the base stations is collapsed, highlighted with a red circle, due to a natural disaster. The UE tries to reach the nearest base station (BS) by increasing the transmitted power. This flood of traffic from the out-of-coverage area of the network results in a congestion in other areas of the network due to the scarce resources available at the BS. In this situation, we prioritize the radio resources for D2D communication among first responders, which are shown by a red circle in Figure 3.

5. Results and Discussions

For simulations, we use NS-3 [13] network simulator extended with LTE functionality from LENA project. The
distance among UE, participating in the direct D2D communication, ranges from 10 m to 50 m, while their distance from the LTE BS varies from 250 m to 500 m. UE transfers a 20 MB content to the rest of the UE in the group. The UE is free to move within the aforementioned distance range. We assume that all UE of a group is identical in terms of communication and processing capabilities.

Figure 4 shows the average latency of a D2D network in sharing a 20 MB content within a group by utilizing the aforementioned slicing and resource allocation policies between VNs. It is important to note that, in order to calculate the average network latency, we measure the latencies of individual UE from the source and average it with the total amount of destination UE in a group.

It can be observed from the figure that the network latency in our scenario depends not only on the distance among UE but also on the mobility of UE. Moreover, latency is directly proportional to the average distance among UE and the amount of UE in a group.

We also measure the task completion time to offload computations to a different D2D group. In this scenario, we divide a computational extensive application into smaller tasks and offload them to the UE of a D2D group to compute the result. We change the size of the group as shown in Figure 5. It can be observed from the figure that, for data sizes less than 5 MB, the task completion time remains independent of the amount of D2D UE in a group. On the other hand, for larger data sizes, an increase in the amount of UE in D2D group makes offloading more efficient in terms of task completion time. However, after a certain threshold, further increase in the amount of UE results in a slight increase in the task execution time. This upper bound of the amount of UE in a D2D group is due to the fact that having too much UE in a group increases the communication and task slicing overhead, which contributes to an increase in the task execution time.

6. Conclusions

In this work, we presented an architecture combining stochastic geometry, resource slicing, and device-to-device communication to guarantee reliability and low latency for public safety networks. Service type and network traffic situation-aware resource slicing increase the flexibility in reconfigurability of the future cellular networks, thus enabling diverse services to be handled over a single infrastructure lowering the capital and operational expenditures to mobile operators. Furthermore, the proposed resource slicing algorithm is tailored to give more priority to first
responders, improving their communication in public safety scenarios by efficiently allocating the highly scarce radio resources. Moreover, priority-based resource slicing enables reliable communication for first responders.

Conflicts of Interest

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

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1617–1655, 2016.
[2] NetWorld2020, "5G: Challenges, research priorities, and recommendations," pp. 1–45, September 2014, https://networld2020.eu/wp-content/uploads/2015/01/Joint-Whitepaper-V12-clean-after-consultation.pdf.
[3] A. A. Gebremariam, M. Chowdhury, A. Goldsmith, and F. Granelli, "Resource pooling via dynamic spectrum-level slicing across heterogeneous networks," in Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 818–823, Las Vegas, Nev, USA, January 2017.
[4] M. Usman, Energy efficiency and privacy in device-to-device communication [Ph.D. thesis], University of Trento, 2017.
[5] 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.
[6] A. Kumbhar, F. Kooohifar, 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.
[7] Y. Shibata, N. Uchida, and N. Shiratori, "Analysis of and proposal for a disaster information network from experience of the Great East Japan Earthquake," IEEE Communications Magazine, vol. 52, no. 3, pp. 44–50, 2014.
[8] 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.
[9] 5GPPP, "5GPPP use cases and performance evaluation models," pp. 1–39, April 2016, https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-use-cases-and-performance-evaluation-modeling_v1.0.pdf.
[10] W. Lu and M. Di Renzo, "Stochastic geometry modeling of cellular networks," in Proceedings of the the 18th ACM International Conference, pp. 179–188, Cancun, Mexico, November 2015.
[11] M. Haenggi, Stochastic Geometry for Wireless Networks, Cambridge University Press, 2013.
[12] 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.
[13] "NS-3, Network Simulator," https://www.nsnam.org/.
