

Received September 9, 2021, accepted November 26, 2021, date of publication December 3, 2021, date of current version December 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132427

On Optimizing Network Function Placement for Multicast Group Call Service Provision in LTE IOPS Networks

JORGE ORTÍN¹, JOSÉ RAMÓN GÁLLEGO², ÁNGELA HERNÁNDEZ-SOLANA², AND MARÍA CANALES¹

¹Centro Universitario de la Defensa, 50090 Zaragoza, Spain

²Aragón Institute of Engineering Research, Universidad de Zaragoza, 50018 Zaragoza, Spain

Corresponding author: Jorge Ortín (jortin@unizar.es)

This work was supported in part by the Spanish Ministry of Science under Grant RTI2018-099063-B-I00, in part by the Gobierno de Aragón through Research Group under Grant T31_20R, and in part by the European Social Fund (ESF).

ABSTRACT The adoption of commercial wireless technologies such as 4G/5G in public safety networks allows for new broadband services that could not be offered in this type of networks. Nevertheless, public safety networks are mission-critical systems that require the introduction of specific services in commercial wireless networks before this adoption. For instance, public safety networks require push-to-talk group communications or the possibility of operating in isolated mode, named by 3GPP the Isolated E-UTRAN Operation for Public Safety (IOPS). These requirements pose specific challenges that must be addressed, such as the placement of a local Evolved Packet Core (EPC) when a set of evolved Node B (eNB) must work in IOPS mode. In this paper, we analyze the implications of push-to-talk group communications and multicast in the local EPC placement problem for IOPS networks and propose an optimization algorithm that finds the best eNB where the local EPC should be co-located. Given the complexity of the problem, we also propose a heuristic algorithm to determine the optimal placement of the local EPC. Simulation results show the ability of our approach to adapt to different types of traffic with heterogeneous spatial distributions.

INDEX TERMS IOPS, LTE, local EPC placement, optimization, multicast.

I. INTRODUCTION

For a long time, there have been two separate technologies for providing wide-area wireless communications: commercial cellular networks (standards 2G/3G/4G of the 3GPP) and dedicated public safety (PS) systems, such as the narrowband systems Terrestrial Trunked Radio (TETRA) and Tetrapol (TETRA for police) in Europe, and Project 25 (P25) in North America. Similar to commercial networks, where Long Term Evolution-Advanced (LTE-A) has become the paradigm of broadband wireless networks, there is a desire for PS to accommodate high-bandwidth data and services. Hence, the transition from the current narrowband to broadband communication becomes the overall objective in PS networks as well. In this context, instead of developing new dedicated PS broadband solutions, integrating PS communications with commercial mobile broadband standards is more beneficial.

The associate editor coordinating the review of this manuscript and approving it for publication was Amjad Ali.

In fact, LTE-A as well as the recent and future releases of 3GPP (LTE-pro and 5G) have been assessed and accepted by multiple authorities and governments as a reference technology for the evolution of PS systems to offer broadband services. The main advantages are: the ability to support voice, video, data and new and enhanced multimedia services simultaneously; network redundancy and reliability; flexible adaptation to multiple frequency bands and spectrum bandwidths; low latency and connection setups; robust and highly reliable transmission; high transmission rates; Quality of Service (QoS) differentiation; real-time positioning and tracking; interworking and integration with existing legacy technologies and applications; and low cost of infrastructure. Furthermore, defined as a specific type of private LTE, the choice of LTE can also be leveraged to unify existing disparate PS networks. In addition, a common technology for both commercial and dedicated PS networks offers many advantages. For instance, sharing network resources reduces costs and deployments.

However, beyond the advantages of LTE, PS networks are mission critical (MC) systems that require satisfying strict requirements in terms of network coverage, system accessibility and reliability, QoS, end-to-end performance and security which commercial systems normally do not require. Service availability and reliable performance must be ensured even in emergency situations, in areas where fixed communication infrastructure may not exist, is destroyed (e.g., due to disaster), or cannot provide sufficient capacity. Thus, the MC aspect of PS communications adds unique demands compared to commercial services. These demands refer to the support of group communications, device-to-device communications, direct mode communications (also known as proximity services, ProSe), MC push-to-talk (PTT) voice, video, and data (MCPTT, MCVideo, MCDATA), end-to-end security, and isolated mode operation. 3GPP began addressing these requirements as part of the evolution of LTE.

The first document concerning PS was published in 3GPP Release 11 [1]. The evolution of MC services includes the specification of Group Communications System Enablers (GCSE) [2] and ProSe [3] in Release 12. GCSE allows efficient distribution of the same content (voice, data, video) to multiple users (User Equipment - UE) of a group at the same time, while ProSe allows UEs in proximity to discover each other and communicate. MCPTT [4], [5], Isolated E-UTRAN Operation for Public Safety (IOPS) [6] and enhanced ProSe (eProSe) [7] were included in Release 13. When an MCPTT communication is established, only one UE can talk by pressing a button while the rest listen. IOPS mode operation [6], [8] allows the maintenance of local IP connectivity and PS services to UEs when the LTE access network (E-UTRAN) has lost backhaul connectivity to the macro Evolved Packet Core (EPC). E-UTRAN encompasses all base stations (named as evolved Node B or eNB), while the macro EPC is the core network that communicates with external packet data networks (PDN) and handles both data flows and signaling. IOPS operation is also required when the backhaul is not fully functional or does not exist. Specifications of MCVideo [9] and MCDATA [10] and enhancements for MC services have been included in Release 14. Interworking with narrowband PS technologies (e.g., TETRA) and studies into future railway mobility communications systems were part of Release 15. Releases 16 and 17 include items of interest to PS in 5G New Radio, while MCPTT on IOPS will not be standardized until Release 17 [11].

This work focuses on IOPS mode operation. To provide MC services (including MCPTT, group calls and dispatching services), isolated networks, consisting of one or more eNBs operating in IOPS mode, must have access to a local EPC. The local EPC instance should locally replicate the behavior of the macro EPC, providing the same basic functionalities. Nevertheless, local EPC dimensioning and optimum placement involve several open issues. Concerning the number of local EPCs, the standard recommends choosing only one local EPC to which all the eNBs of an IOPS network are connected. Nevertheless, in an IOPS area, multiple active

local EPCs instances may be considered if required. In any case, even if only one instance of the local EPC is active, this does not prevent the coexistence of multiple optional local EPCs co-located in different eNBs in the IOPS network. In this case, the questions to be answered are how many pre-existing local EPCs must be placed in the network, their preferred location, and the most important issue, which one of the pre-existing local EPCs must be activated. Local EPC placement and selection depends on multiple factors: data and signaling traffic distribution and requirements, types of service provision (unicast, multicast), capacity restrictions of local backhaul links between eNBs, and topology (an eNB may require direct links or multi-hop to reach the active local EPC). Studies on this issue are scarce. In this paper, the active local EPC placement problem is addressed as an Integer Linear Programming (ILP) problem with specific restrictions concerning air and link capacities and details of MC services such as MCPTT group calls.

The paper is organized as follows: Section II summarizes the basis and challenges in the IOPS operation mode and MC services, focusing on group MC group services. Section III reviews the related work on the function placement problem in wireless networks and specifically of IOPS entities to clearly state the contributions of our work. Section IV describes the network model, the optimization problem to compute the best local EPC location, a complexity analysis of the optimization problem and an alternative heuristic algorithm of lower complexity. The results are discussed in Section V, followed by the concluding remarks in Section VI.

II. IOPS AND PUBLIC SAFETY SERVICES

A. IOPS BASIS AND CHALLENGES

IOPS operation for PS allows UEs to maintain a level of communication, via one isolated eNB or a set of interconnected isolated eNBs (which form an isolated E-UTRAN), when they have lost the backhaul link to the core LTE network infrastructure [6], [8]. Isolated E-UTRAN operation requires eNBs to be IOPS-capable eNBs. That is, when the backhaul between the eNBs and the macro EPC decreases, the IOPS-capable eNBs can detect the loss of the backhaul, initiate the IOPS mode operation based on local policies and start providing local IP connectivity and PS services to their UEs through a local EPC instance. The local EPC, which is co-located with the eNB or reachable through other eNBs in the isolated E-UTRAN, must include at least the main entities of the macro EPC and means to locally deliver security/access control. Fig. 1 shows the network architecture for the normal vs. IOPS operation modes. In the control plane, it must include the Mobility Management Entity (MME), which is responsible for UE mobility management (idle tracking and handover), paging, selection of the Serving Gateway (S-GW), bearer activation/deactivation processes and UE authentication with the Home Subscriber Server (HSS). Thus, the Local EPC needs to perform security and authentication and provide necessary HSS functions to support UE access to the Local

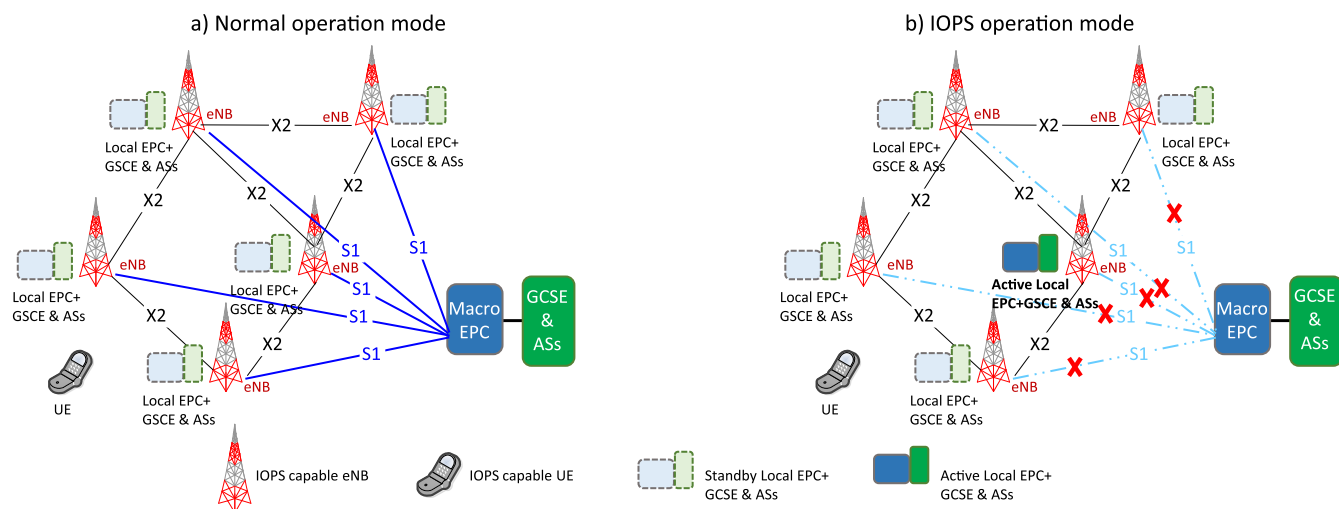


FIGURE 1. The normal vs. IOPS operation modes.

EPC, via subscription data and authorization. Note that in IOPS mode, the eNBs start advertising a Public Land Mobile Network (PLMN-Id) dedicated to IOPS and only authorized UEs can access this PLMN. On the other hand, if the eNB cannot reach a local EPC instance, the eNB must enter a state that prevents UE from selecting the cells under its control. In the data plane, the Local EPC includes two main entities: the S-GW, whose function is the routing and forwarding of data packets and the Packet Data Network Gateway (PDN-GW or P-GW), which is the connection point of the UE with external PDNs and Application Services (AS) servers.

Isolated E-UTRAN operations include not only fixed eNBs with lost connectivity to the macro EPC, but also standalone Nomadic eNBs (NeNB). NeNBs are deployable systems, including a transportable local EPC that assists PS services in areas without infrastructure. The IOPS and nomadic deployments are conceptually similar. Thus, an isolated E-UTRAN with NeNBs exhibits a behavior similar to that of an isolated E-UTRAN with either eNBs or mixed NeNBs and eNBs. In addition, IOPS addresses not only the absence of backhaul, but also limited backhaul bandwidth. In this case, IOPS considers two possible scenarios according to the transmitted data in the limited backhaul: i) only signaling is transmitted to the macro EPC while user data is routed locally or ii) both signaling and a limited amount of user data are transmitted to the macro EPC. In this case, user data has no guarantee of service when it is routed through the backhaul. In this study, we focus on scenarios where there is no backhaul, so the deployment of a local EPC is mandatory, and within these scenarios, on the optimum local EPC placement.

As stated in the previous section, an IOPS network may comprise a single isolated eNB (which may be co-located with a local EPC instance) or more eNBs. In the last case, all eNBs establish S1-MME paths to the local MME of the same local EPC instance, which is co-located with one of the eNBs. This does not prevent several isolated E-UTRAN,

each one with its local EPC instance, may be combined to define a larger IOPS network. However, in this case, the tracking area codes broadcasted by the cells of the eNBs connected to different local EPCs must be different to ensure network identification and the required UE mobility behavior. In addition, mobility management must include Inter-IOPS network mobility.

In this work, we focus on the case with only one local EPC; thus, only one local MME is used to ensure the connectivity of UEs in an IOPS region composed of several eNBs. In this context, the aim is to quantify the satisfaction of UE's service and traffic demands. Because all data and signaling traffic must be transferred to the single active local EPC, optimum dimensioning involves not only calculating the local EPC capacity to handle data and signaling, but also its optimum location. As all the eNBs in the IOPS network must reach the local EPC, the problem requires explicit consideration of the features and capacity of the inter-eNB connectivity, and an accurate estimation of the signaling and data transport needs.

Moreover, it should be noted that one or more candidate Local EPC instances can exist in the coverage area of the IOPS network. In fact, IOPS-capable eNBs can be provisioned with the IP endpoint of a preferred local EPC MME instance and the IP endpoints of one or more alternative local EPC MME instances in the area. As shown in Fig. 1, one, several, or all eNBs may be pre-provisioned with co-located standby local EPC instances. Nevertheless, when several eNBs joint to form an IOPS network they must share the same active MME instance. Alternative instances are used if a path cannot be established with the preferred local EPC. In this case, unreachable eNBs may configure a separate IOPS network. On the other hand, the preferred local EPC election may be performed in a semi-static fashion when the network is operating in normal mode to optimize data and signaling provision when some eNBs in a region enter in IOPS mode. Thus, placement of the selected local EPC, each

time a group of eNBs enters the IOPS mode is a relevant issue and depends on the UE and traffic distribution.

Multiple S-GW instances linked to an MME could be considered in the IOPS configuration, because the location of the S-GW to which a UE is attached determines the routing path of data traffic in the backhaul. A smaller backhaul bandwidth could be consumed if traffic passed through an S-GW that was on the shortest path between the eNBs. However, the location of the P-GW must also be considered when MC services are deployed in the IOPS network, because the P-GW is involved in MCPTT group call data paths as we will show in Section 2.2. Therefore, in the end, (i) having only one instance of the S-GW/P-GW, (ii) distributing the S-GW instances on all eNBs, separated from the P-GW, or (iii) selecting a subset of eNBs to co-locate several S-GW instances, have a limited impact on the consumption of backhaul bandwidth because all the user data paths must go through the P-GW. In addition, the co-location of the S-GW instance in a separate eNB compared to the MME could be also considered, but when the MME and the S-GW are co-located, the signaling traffic between them is not routed on the backhaul, economizing bandwidth.

Thus, we consider that all the entities (MME/S-GW/P-GW) are physically co-located in a single local EPC placement as this is the most efficient approach when group MC services are supported. Additionally, and as we will see in the next section, apart from the MME, S-GW and P-GW, a local MCPTT system must be deployed at the IOPS network to provide MCPTT services, including MCPTT group calls.

B. PUBLIC SAFETY SERVICES

Voice communications requirements in PS networks include support of common Full Duplex (FD) voice like that used in commercial wireless networks as well as PTT, which is the current standard for voice communications in PS networks. In this case, UEs cannot speak simultaneously but individually in specific time slots. The speaker presses the PTT button on the radio to talk to other units and releases it to return to the listen mode. In addition, group calls are required in the PTT paradigm.

Being of vital importance to PS, group calls have been added by 3GPP within the MCPTT services defined in Release 13. In fact, many technical specifications of Release 12 and 13 focus on group calls, including the general core specification GCSE [12] and MCPTT [13]. MCPTT group calls are intended to support communication between multiple UEs (group), with the ability to grant access to speak. When multiple requests occur, the mechanisms to determine which UE request is accepted, rejected or queued depend on a number of features (including priorities of UEs in contention) that are out of the scope of this paper. Nevertheless, functional architecture requirements to support these services and bearer transport need to be clearly stated. MCPTT group calls comprise a set of architectural concepts from core LTE, MCPTT, Common Services Core (CSC) for Mission Critical Communication (MCC) services and GCSE. Fig. 2 shows the

entities required to support these services. For instance, the MCPTT Application Server (MCPTT AS), which is responsible for delivering a half duplex group voice communication service, includes floor control and media distribution function. Configuration, group management server, etc. at the CSC provides support to all MCC services. This framework, originally defined for voice, was expanded with the inclusion of MCData and MCVideo in Release 14 and beyond [9], [10]. Regarding MCVideo group calls, although MCVideo service includes news operation modes added to MCPTT, we focus on MCVideo group calls based on PTT operation mode.

Regarding the data delivery of a group call, Release 12 considers conventional physical channels in LTE for data delivery: i) the Physical Downlink Shared Channel (PDSCH), which is commonly used for normal unicast data, and ii) the Physical Multicast Channel (PMCH), which is designed for evolved Multimedia Broadcasting and Multicasting Service (eMBMS) [12], [14]–[16]. Both cases are evaluated in this study. In the first stage, still considered by many operators, each group UE is connected separately via unicast Evolved Packet System (EPS) bearers in uplink (UL) and downlink (DL) to the MCPTT and GCS ASs (see UEs attached to eNB1 and the speaker UE of eNB2 in Fig. 2). In this case, the group data in DL should be replicated according to the number of UEs per group. Given that the EPS bearer concatenates the radio bearer and the core network bearer to the P-GW, radio resources linked to dedicated radio bearers need to be allocated by the eNB separately for each UE in the group, but also core resources. This is a waste of resources and a bottleneck to satisfy scalability requirements [17], although the problem must be analyzed in more detail. If the air interface is the main bottleneck, an advantage of the unicast bearer is that the eNB can apply advanced link adaptation schemes on the air interface, such as adaptive Modulation and Coding Schemes (MCS), HARQ, or MIMO, which significantly increases the spectral efficiency and reliability. This may partially or fully offset the losses due to the duplication of resources and improve scalability. In addition, the end-to-end delay can be shortened as the data goes to the standard core network (eNB ↔ S-GW ↔ P-GW ↔ GCS AS).

However, owing to the standardization of MBMS/eMBMS multicast bearers, data can be delivered via shared network resources to multiple UEs [15]. Multicast can be considered in two ways: Single-Cell Point-to-Multipoint (SC-PTM) bearer [16] and as a part of the Multicast-Broadcast Single-Frequency Network (MBSFN) topology. In both cases, new entities were required. The Broadcast/Multicast Service Center (BM-SC) is the entry point of the MBMS content provider. It allows the MCPTT AS to stream media through multicast bearers over the LTE Radio Access Network (RAN) and controls the overall service (it manages the MBMS sessions -start, update and stop service announcements- and provides security to the sessions). The MBMS Gateway (MBMS-GW) distributes the MBMS data to each eNB using IP multicast. Finally, the Multi-cell/multicast Coordination Entity (MCE) handles the radio resource allocation and the coordination

spectral efficiency and flexibility due to PDSCH election on the other.

In general, we can conclude that there is a trade-off between the advantages and limitations of unicast and multicast options depending on the size of the group and the distribution of UEs. Unicast can be a good option for serving small groups or for operators without multicast active operations. MBSFN is suitable for large groups and SC-PTM is a compromise option between them. In fact, MCE (located in the eNBs or centralized to control several eNBs) not only handles radio resource allocation but also the coordination of MBSFN and SC-PTM. It decides whether to set up the eMBMS bearer depending on the resources available in the cells. Therefore, we can conclude that all options should be considered. Thus, in this paper, we include the impact of both unicast and multicast for the provision of MCPTT group calls in our optimization problem.

To support multicast, the local EPC must add the entities of Local BM-SC, MGMS-GW, and MCE in addition to the MME, S-GW, and P-GW. The IOPS MC subsystem, which includes application functions and enabling capabilities to support MC services such as MCPTT group calls, must be added as well. In this work, the IOPS network comprises a local EPC/IOPS MC Services instance and two or more eNBs, one of which co-located with the local EPC/IOPS MC Services instance (see Fig. 2(b)). Note that the IOPS MC subsystem is on standby during normal operation. According to [11], the IOPS connectivity client and the MCPTT client (to allow the UE to be discovered and registered in the IOPS MC subsystem) must be included in the UE, whereas the IOPS MC connectivity and IOPS distribution functions provide support to MC services (register and discover UEs) and handling of IP packets containing the MC service application data received from a UE in IOPS mode.

Without loss of generality and to simplify the analysis, we can address the local EPC placement problem without explicitly by considering the details about interference and scheduling effects at the air interface to set air resource requirements. Instead, a conservative estimation of the resource requirements is used. Thus, the main effects of multicast have been considered together for the SC-PTM of MBMS implementations, as both share the functional architecture of MBMS and IP multicast transmission at the core. In each eNB, at the air interface, both implementations require a unique and common DL resource allocation for voice media transmission for each group flow, while unicast requires as many resource allocations as UEs in the group.

Note that in addition to voice/video media transmission over a dedicated bearer, default bearers are always present ensuring always-on IP connectivity and signaling between UEs and MC entities. Signaling can be mapped to a different unicast bearer for the UE, as shown in Fig. 2(c). In addition, signaling radio bearers 0, 1, and 2 are always present in each UE to support the transmission of radio resource control and non-access-stratum signaling messages between the UE and

core LTE entities (for instance, linked to attach procedures, location updates, mobility management, etc.).

III. RELATED WORK AND MAIN CONTRIBUTIONS

In wireless networks, the optimum placement problem has been addressed in the literature linked to several entities, such as the gateway or sink placement problem in multi-hop wireless networks and wireless mesh networks. In addition, recent advances in the fields of network softwarization and virtualization have led to the proposal of multiple works addressing network function placement problems [18]–[20] that mainly focused on cloud and edge computing. In any case, the differences both due to the network constraints in the wireless environment, and the specific system requirements, such as MC services or multicast transport, make these problems not directly comparable with this work. Specific studies related to optimizing IOPS settings are scarce, mainly limited to the works performed by Queis *et al.* [21]–[23].

Namely, the work in [21] tackles the local EPC placement problem from a centralized point of view. That is, in an isolated network composed of various eNBs, the local EPC must be co-located in only one of the eNBs. When considering the local EPC as an endpoint, the goal is to determine the eNB where the local EPC should be co-located to maximize its capacity to receive and forward data and signaling from the eNBs of the network. The intuition is that the local EPC should be a “central” node. To do this, they defined a centrality metric that measures the capacity of a node to maximize receiving traffic. According to the authors, many previous node-centrality metrics, such as degree centrality [24], weighted degree [24], closeness [25], and betweenness centrality [26], are very limited because they do not include full information on the link capacities of the network. Because of this, they proposed a new metric called flow centrality. This metric measures the maximum uniform traffic that nodes in the network can send simultaneously to the central node where the local EPC is co-located given link capacity constraints both in the inter-eNB links and in the air interface. Traffic between the eNB and local EPC is routed directly or by multi-hop interconnected eNBs.

The main limitation of this work is that all eNBs are assumed to generate the same amount of traffic toward the local EPC, without considering the traffic generation patterns of the UEs (type of service, rate requirements or sources and destinations). That is, the objective function is the maximum amount of traffic that the local EPC can receive from the eNBs, as a single destination. The same authors in [22] extended this analysis by showing that in some network graphs, such as path graphs and balanced trees, it is possible to calculate the flow centrality of a node using analytical expressions instead of solving a linear optimization problem.

Data traffic and signaling impact in the backhaul (links interconnecting eNBs) of the IOPS network are considered by the authors in [23]. They proposed an optimization problem with the aim of minimizing the backhaul bandwidth consumption of user and signaling information exchanged

between the eNBs and the MME and S-GWs. They assumed that only one MME is needed in the network and considered three different options for the deployment of the S-GWs: i) a unique instance of the S-GW, ii) an instance of the S-GW co-located with every eNB and iii) the optimal number and distributions of S-GW instances. In all cases, they found the best location for the MME.

The main results of this study are as follows: First, in case it is mandatory to have just one S-GW in the network, the best option is co-locating the MME and the S-GW in the same place to avoid signaling traffic between them. Second, having an S-GW co-located with every eNB performs similarly to the optimally distributed option. This result is also logical, as they assume that the data flows between UEs can be established directly without passing through a central P-GW, and the impact of the control signaling is almost negligible compared to the data distribution. Nevertheless, in PS scenarios supporting MCPTT group calls, the data flow must go through the MCPTT AS (where media distribution and floor control server are present) via the S-GW and P-GW before being forwarded to the destination. This means that in this scenario data flows between S-GW/PDN-GW and AS in addition to data flows between eNBs and S-GW must be included in the analysis. As a result, the co-location of the S-GW with the rest of the local EPC functions is the preferred option. Finally, in none of the described works regarding entity placement in IOPS networks, the capability of having PTT or multicast transmission has been considered.

With regard to the optimum placement problem in other wireless networks, several works deal with the gateway location problem in multi-hop and mesh wireless networks through mathematical optimization with linear programming or from a heuristic perspective [27]–[30]. The main objectives are similar to the local EPC placement. The goal is to minimize the number of gateways and optimize their placement under different QoS constraints (throughput maximization, delay requirements, etc.). However, the formulation of this problem is quite different. The functional architecture requirements and distribution of entities in the data and control planes are different. As stated in the previous section, in our case, centralized placement is preferred, since all the data flows in the network may require going through a certain function. In addition, other typical constraints considered in multi-hop wireless networks, such as energy consumption in wireless sensor networks (WSN), are not very relevant in the location of the local EPC. Another difference with WSN is that resource allocation in the inter-eNB links is orthogonal, and even at the air interface, this assumption can be made under some considerations, owing to the effects of the centralized scheduling in the eNB. In WSNs, these assumptions are not possible, which translates into different optimization problems.

In this context, the contributions of the paper are:

- (i) We study the optimum placement of the preferred local EPC/IOPS MC services instance in the IOPS mode (alternative standby local EPCs can coexist) as well as the paths to reach this preferred local EPC for each data flow in the network. The aim is to maximize the number of connections that can be simultaneously supported on the network.
- (ii) We define the problem as an ILP problem, explicitly considering the implications of specific services for PS: MCPTT and full duplex services and unicast vs. multicast group call MC voice and video services. Air and core link capacity constraints and resource demands of traffic sources are calculated more closely to a real MC deployment, according to the LTE standard limitations.
- (iii) We theoretically analyse the complexity of the algorithm, demonstrating that it is NP-hard. Because this can lead to scalability problems, we also propose a heuristic algorithm for the local EPC placement based on the relaxation of the original problem to obtain sub-optimal results in reduced computational time.
- (iv) The evaluation was performed in scenarios where mixed traffic of different classes coexisted. The impact of a non-homogeneous distribution of traffic is considered in the selection of the preferred local EPC.

IV. LOCAL EPC PLACEMENT PROBLEM

A. NETWORK MODEL

We consider a network of interconnected IOPS eNBs that have lost backhaul with macro EPC. Although multiple standby local EPC/IOPS MC Services instance (each one co-located with an eNB) are considered, a unique active local EPC/IOPS MC Services instance (for brevity, local EPC from now on) serves the network (the preferred entity). The local EPC provides the same functionality as the macro EPC, adding MBMS entities to support multicast services and IOPS MC and application servers. Inter-eNB links are defined between each eNB. Without loss of generality, the specific technology used in the inter-eNB links was not considered. Nevertheless, to model a scenario with limited inter-eNB capacity, the transmission rates are limited to those achievable if out-of-band LTE carriers (different from the one used in the air interface) were used in the inter-eNB connections. Data and signaling traffic between each eNB and the local EPC are transmitted directly or through interconnected eNBs in a multi-hop fashion. Intercell interference is not explicitly considered at the wireless interface. Instead, the total capacity available at the air interface is estimated considering a conservative approach, assuming the use of a mean MCS (intermediate between the most robust and the highest rate).

Qualitatively, the optimization problem that we aim to solve can be defined as follows: to maximize the weighted number of FD calls and group calls that can be established in the network subject to the following constraints: i) system constraints, that is, each UE in an FD call or group call must be attached to an eNB and must reach a local EPC/IOPS MC. In addition, there must be a single local EPC/IOPS MC on the network through which all FD calls and group calls pass;

ii) routing constraints to set paths between different eNBs and the local EPC/IOPS MC; iii) bandwidth constraints to ensure that the transmission resources allocated to the different FD calls / group calls at the radio interface of each eNB and the backhaul links do not exceed the available capacity.

B. OPTIMIZATION FRAMEWORK

1) OBJECTIVE FUNCTION

Table 1 summarizes the notation used in this section. Let \mathcal{B} be the set of eNBs working in isolated mode, \mathcal{U} the set of UEs under the coverage of \mathcal{B} , and \mathcal{T} and \mathcal{G} be the set of FD and group calls to be established between UEs in \mathcal{U} respectively. To simplify the notation, in the following, we use the index k to go through the different elements of \mathcal{T} , m for the elements of \mathcal{G} , i for the elements of \mathcal{U} and j for the elements of \mathcal{B} . For each $m \in \mathcal{G}$, we define the set \mathcal{U}_m as the subset of UEs of \mathcal{U} that want to join the group call m . Likewise, for each $k \in \mathcal{T}$, we use the terms $U_k^{(1)}$ and $U_k^{(2)}$ to denote the two UEs that communicate through FD call k .

Each UE $i \in \mathcal{U}$ is covered by a subset of eNBs of \mathcal{B} , which we denote by \mathcal{B}_i . Additionally, each eNB is also connected to a subset of other eNBs in \mathcal{B} , which we denote by \mathcal{B}_j . We assume that the eNBs in \mathcal{B} form a connected network to ensure the existence of a solution of the optimization problem (if not, we would have more than one network working in isolated mode).

Further, let t_k be a binary variable indicating whether FD call $k \in \mathcal{T}$ is established in the network and $g_{m,i}$ is a binary variable indicating whether UE i of group call $m \in \mathcal{G}$ has joined the group call m . The objective function aims to maximize the number of FD calls / UEs in group calls that can be served in the network while minimizing the use of resources in the backhaul:

$$\max \sum_{k \in \mathcal{T}} p_k t_k + \sum_{m \in \mathcal{G}} \sum_{i \in \mathcal{U}_m} p_{m,i} g_{m,i} - \delta r_{\text{bh}} \quad (1)$$

where p_k and $p_{m,i}$ are respectively the preference weights of the different FD calls / UEs in the group calls, δ is a penalty factor associated with the use of the backhaul such that $\delta \ll p_k$ and $\delta \ll p_{m,i}$ and r_{bh} is a real variable indicating all the resources used in the backhaul.

2) SYSTEM CONSTRAINTS

First, for an FD call to be established (or a UE to join a group call), we need the UEs of the FD call (or the group call) to be connected to an eNB. We also need that they are registered to the local EPC of the network. If we define the binary variables $b_{i,j}^k$ ($b_{i,j}^m$) to denote that UE i of FD call k (group call m) is connected to eNB j and the variables $h_{i,j}^k$ ($h_{i,j}^m$) to denote that the active local EPC is co-located with eNB j and UE i of FD call k (group call m) is also registered to it, we have the

following set of constraints

$$t_k = \sum_{j \in \mathcal{B}_i} b_{i,j}^k \quad \forall k \in \mathcal{T}, i = U_k^{(1)}, U_k^{(2)} \quad (2a)$$

$$t_k = \sum_{j \in \mathcal{B}} h_{i,j}^k \quad \forall k \in \mathcal{T}, i = U_k^{(1)}, U_k^{(2)} \quad (2b)$$

(2) indicates that for FD call k to be established, both its UEs must be connected to an eNB (2a) and registered in the local EPC of the network (2b). Note that as all variables are binary, the previous restrictions ensure that a UE i is connected to just one eNB as well. The following set of constraints are equivalent to those in (2), but for the UEs of group calls

$$g_{m,i} = \sum_{j \in \mathcal{B}_i} b_{i,j}^m \quad \forall m \in \mathcal{G}, i \in \mathcal{U}_m \quad (3a)$$

$$g_{m,i} = \sum_{j \in \mathcal{B}} h_{i,j}^m \quad \forall m \in \mathcal{G}, i \in \mathcal{U}_m \quad (3b)$$

Now, we have to guarantee that just one local EPC is deployed in the network. To do so, let h_j be a binary variable indicating whether the local EPC is co-located with eNB j . First, we need the set of constraints (4) to ensure that the local EPC is co-located with an eNB if it is at least co-located with this eNB for one UE

$$h_j \geq h_{i,j}^k \quad \forall j \in \mathcal{B}, k \in \mathcal{T} \text{ and } i = U_k^{(1)}, U_k^{(2)} \quad (4a)$$

$$h_j \geq h_{i,j}^m \quad \forall j \in \mathcal{B}, k \in \mathcal{G} \text{ and } i \in \mathcal{U}_m \quad (4b)$$

Once we have the variables h_j , the following constraint guarantees that the active local EPC is co-located with just one eNB

$$\sum_{j \in \mathcal{B}} h_j = 1 \quad (5)$$

3) ROUTING CONSTRAINTS

Each established FD call in the network must go from the eNBs to which the UEs are connected to the local EPC and back, so we must guarantee that there is a path from each eNB to the local EPC. To do so, we apply a fluid model to each eNB of the network. Namely, let $f_{j,j'}^{k,B_i \rightarrow E}$ be a binary variable indicating whether FD call k is routed through the link between eNBs j and j' from the eNB of UE i to the local EPC (denoted by E). Then, the fluid model can be conveniently expressed using the set of constraints

$$b_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{k,B_i \rightarrow E} = h_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{k,B_i \rightarrow E} \leq 1 \quad \forall j \in \mathcal{B}_i, k \in \mathcal{T}, i = U_k^{(1)}, U_k^{(2)} \quad (6a)$$

$$\sum_{j' \in \mathcal{E}_j} f_{j',j}^{k,B_i \rightarrow E} = h_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{k,B_i \rightarrow E} \leq 1 \quad \forall j \notin \mathcal{B}_i, k \in \mathcal{T}, i = U_k^{(1)}, U_k^{(2)} \quad (6b)$$

TABLE 1. Summary of notation.

Model Parameters	
\mathcal{B}	Set of eNBs
\mathcal{B}_i	Set of eNBs to which UE i can be connected
\mathcal{E}_j	Set of eNBs directly connected to eNB j
\mathcal{G}	Set of group call to be established into the network
\mathcal{T}	Set of FD calls to be established into the network
\mathcal{U}	Set of UEs
\mathcal{U}_m	Set of UEs of group call m
$U_k^{(1)}, U_k^{(2)}$	The two UEs communicating via FD call k
$c_{d,j}/c_{u,j}$	Available capacity at the DL/UL of an eNB
$c_{j,j'}$	Capacity of the link between eNBs j and j'
p_k	Preference weight of FD call k
$p_{m,i}$	Preference weight of UE i of group call m
$r_{d,k}/r_{u,k}$	Transmission rate required for FD call k in the DL/UL
$r_{d,m}/r_{u,m}$	Transmission rate required for group call m in the DL/UL
r_k	Transmission rate required for FD call k in the backhaul
r_m	Transmission rate required for group call m in the backhaul
δ	Penalty factor for the use of the backhaul
Model Variables	
$b_{i,j}^k$	Binary variable indicating whether UE i of FD call k is connected to eNB j
$b_{i,j}^m$	Binary variable indicating whether UE i of group call m is connected to eNB j
b_j^m	Binary variable indicating whether any UE of group call m is connected to eNB j
$f_{j,j'}^{k,B_i \rightarrow E}$	Binary variable indicating whether FD call k is routed through the link between eNBs j and j' from the eNB of UE i to the local EPC
$f_{j,j'}^{k,E \rightarrow B_i}$	Binary variable indicating whether FD call k is routed through the link between eNBs j and j' from the local EPC to the eNB of UE i
$f_{j,j'}^{m,B_i \rightarrow E}$	Binary variable indicating whether group call m is routed through the link between eNBs j and j' from the eNB of UE i to the local EPC
$f_{j,j'}^{m,B \rightarrow E}$	Binary variable indicating if group call m is routed through the link between eNBs j and j' for the flow from an eNB to which any UE of group call m is connected to the local EPC
$f_{j,j'}^{m,E \rightarrow B_i}$	Binary variable indicating whether group call m is routed through the link between eNBs j and j' from the local EPC to the eNB of UE i
$f_{j,j'}^{m,E \rightarrow B}$	Binary variable indicating whether group call m is routed through the link between eNBs j and j' for the flow from the local EPC to an eNB to which any UE of group call m is connected
$g_{m,i}$	Binary variable indicating whether UE i of group call $m \in \mathcal{G}$ has joint the group call m
$h_{i,j}^k$	Binary variable indicating that the local EPC is co-located with eNB j and UE i of FD call k is also registered to it
$h_{i,j}^m$	Binary variable indicating that the local EPC is co-located with eNB j and UE i of group call m is also registered to it
h_j	Binary variable indicating whether the local EPC is co-located with eNB j
r_{bh}	Real variable indicating all the resources used in the backhaul
t_k	Binary variable indicating whether FD call $k \in \mathcal{T}$ is established

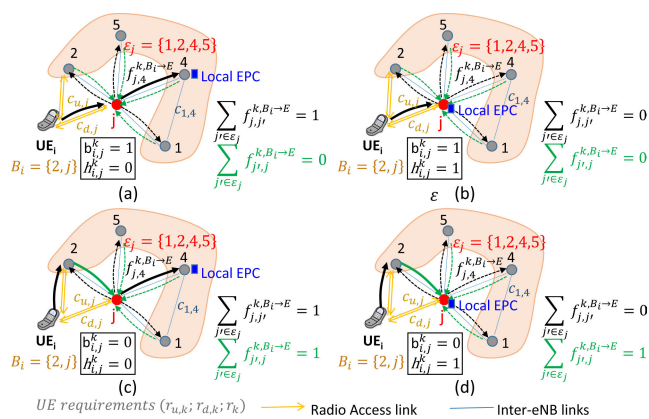


FIGURE 3. Example of the routing constraints in a node: (a) UE i is connected to eNB j , and the local EPC is not co-located with j ; (b) UE i is connected to eNB j , and the local EPC is co-located with j ; (c) UE i is not connected to eNB j , and the local EPC is not co-located with j ; (d) UE i is not connected to eNB j , and the local EPC is co-located with j .

where \mathcal{E}_j indicates the set of eNBs connected directly to eNB j (i.e., there exists a link between eNB j and eNBs in \mathcal{E}_j).

For brevity, each constraint in (6) refers to two constraints. The first constraint corresponds to the flow balance in the eNB (i.e., the two sides of the equality), while the second constraint is a bound for the number of incoming / outgoing routes from each eNB (i.e. the ≤ 1).

These constraints enforce flow conservation at each eNB regardless of whether UE i of FD call k is connected to it, or whether the active local EPC is co-located with that eNB. If it is connected to it, and the local EPC is not co-located with that eNB, then $b_{i,j}^k = 1, h_{i,j}^k = 0$ and from (6) all the incoming links from j will be zero and one outgoing link will be 1 (see the example in Fig. 3(a)). If the local EPC is located with the eNB and UE i is not connected to it, then $b_{i,j}^k = 0, h_{i,j}^k = 1$ and all outgoing links from j will be zero and one incoming link will be 1 (Fig. 3(d)). When the local EPC is not co-located with the eNB or UE i is connected to it, then $b_{i,j}^k = 0, h_{i,j}^k = 0$ and (6) leads to two alternatives: i) the flow is not routed through j , and all the incoming and outgoing links for that flow are zero, or ii) the flow is routed through j and then one incoming and one outgoing link will be 1 (Fig. 3(c)). Finally, if UE i is connected to the eNB and the

local EPC is co-located with that eNB, then there would not be any flow and all the incoming and outgoing links would be zero (Fig. 3(b)).

Following a similar approach, the restrictions for the path back from the local EPC to the eNB to which a UE is connected are

$$h_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{k,E \rightarrow B_i} = b_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{k,E \rightarrow B_i} \leq 1 \quad \forall j \in \mathcal{B}_i, \quad k \in \mathcal{T}, \quad i = U_k^{(1)}, U_k^{(2)} \quad (7a)$$

$$h_{i,j}^k + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{k,E \rightarrow B_i} = \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{k,E \rightarrow B_i} \leq 1 \quad \forall j \notin \mathcal{B}_i, \quad k \in \mathcal{T}, \quad i = U_k^{(1)}, U_k^{(2)} \quad (7b)$$

where $f_{j,j'}^{k,E \rightarrow B_i}$ is a binary variable indicating whether FD call k is routed through the link between eNBs j and j' for the path from the local EPC to the eNB UE i is connected to.

For group calls, the flow conservation constraints are identical to those in (6) and (7), but considering that the flow corresponds to group call m instead of FD call k , that is,

$$b_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{m,B_i \rightarrow E} = h_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{m,B_i \rightarrow E} \leq 1 \quad \forall j \in \mathcal{B}_i, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (8a)$$

$$\sum_{j' \in \mathcal{E}_j} f_{j',j}^{m,B_i \rightarrow E} = h_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{m,B_i \rightarrow E} \leq 1 \quad \forall j \notin \mathcal{B}_i, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (8b)$$

$$h_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{m,E \rightarrow B_i} = b_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{m,E \rightarrow B_i} \leq 1 \quad \forall j \in \mathcal{B}_i, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (9a)$$

$$h_{i,j}^m + \sum_{j' \in \mathcal{E}_j} f_{j',j}^{m,E \rightarrow B_i} = \sum_{j' \in \mathcal{E}_j} f_{j,j'}^{m,E \rightarrow B_i} \leq 1 \quad \forall j \notin \mathcal{B}_i, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (9b)$$

4) BANDWIDTH CONSTRAINTS

Before defining the bandwidth constraints themselves, we must model the impact of PTT and multicast in use of radio resources. As group communications use PTT, there cannot be two UEs of the same group call transmitting simultaneously. This implies that for the UL in the air interface and the backhaul from the eNBs to the local EPC, the global resource allocation is reduced to that required by one UE of the group. To model this, let b_j^m be a binary variable indicating whether any UE of group call m is connected to eNB j and $f_{j,j'}^{m,B \rightarrow E}$ a binary variable denoting if group call m is routed through the link between eNBs j and j' for the flow from any eNB to which any UE of group call m is connected to the local EPC. Then, we can write the following constraints to set the values of these variables

$$b_j^m \geq b_{i,j}^m \quad \forall m \in \mathcal{G}, \quad i \in \mathcal{U}_m, \quad j \in \mathcal{B}_i \quad (10)$$

$$f_{j,j'}^{m,B \rightarrow E} \geq f_{j,j'}^{m,B_i \rightarrow E} \quad \forall j \in \mathcal{B}, \quad j' \in \mathcal{E}_j, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (11)$$

Similarly, when multicast is used, the same resources can be shared for all the UEs of the group call in the DL of the air interface and the backhaul from the local EPC to the eNBs to which the UEs of the group call are connected. If we define the binary variable $f_{j,j'}^{m,E \rightarrow B}$ to indicate if group call m is routed through the link between eNBs j and j' in the way back from the local EPC to an eNB to which any UE of group call m is connected, then we have

$$f_{j,j'}^{m,E \rightarrow B} \geq f_{j,j'}^{m,E \rightarrow B_i} \quad \forall j \in \mathcal{B}, \quad j' \in \mathcal{E}_j, \quad m \in \mathcal{G}, \quad i \in \mathcal{U}_m \quad (12)$$

With this, the following set of constraints models that the transmission resources used by the different FD calls and group calls cannot exceed the available capacity in the UL of the eNBs

$$\sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_{u,k} b_{i,j}^k + \sum_{m \in \mathcal{G}} r_{u,m} b_j^m \leq c_{u,j} \quad \forall j \in \mathcal{B} \quad (13)$$

where $r_{u,k}$, $r_{u,m}$ represent the transmission rate required for FD call k or group call m in the UL and $c_{u,j}$ the available capacity in the UL of eNB j . For the DL, the constraints depend on whether or not we use multicast. When multicast is used, the constraints are similar to those of the UL

$$\sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_{d,k} b_{i,j}^k + \sum_{m \in \mathcal{G}} r_{d,m} b_j^m \leq c_{d,j} \quad \forall j \in \mathcal{B} \quad (14)$$

with $r_{d,k}$, $r_{d,m}$, $c_{d,j}$ corresponding to the same terms as $r_{u,k}$, $r_{u,m}$, $c_{u,j}$ but for DL instead of UL. When multicast is not used, each UE in the group call needs their own transmission resources, so the previous restrictions turns into

$$\sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_{d,k} b_{i,j}^k + \sum_{m \in \mathcal{G}} \sum_{i \in \mathcal{U}_m} r_{d,m} b_{i,j}^m \leq c_{d,j} \quad \forall j \in \mathcal{B} \quad (15)$$

Finally, we have the following bandwidth constraints for the links between eNBs when multicast is not used

$$\sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_k (f_{j,j'}^{k,B_i \rightarrow E} + f_{j,j'}^{k,E \rightarrow B_i}) + \sum_{m \in \mathcal{G}} r_m f_{j,j'}^{m,B \rightarrow E} + \sum_{m \in \mathcal{G}} \sum_{i \in \mathcal{U}_m} r_m f_{j,j'}^{m,E \rightarrow B_i} \leq c_{j,j'} \quad \forall j \in \mathcal{B}, \quad j' \in \mathcal{E}_j \quad (16)$$

while for multicast the constraints are

$$\sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_k (f_{j,j'}^{k,B_i \rightarrow E} + f_{j,j'}^{k,E \rightarrow B_i}) + \sum_{m \in \mathcal{G}} r_m (f_{j,j'}^{m,B \rightarrow E} + f_{j,j'}^{m,E \rightarrow B}) \leq c_{j,j'} \quad \forall j \in \mathcal{B}, \quad j' \in \mathcal{E}_j \quad (17)$$

where $c_{j,j'}$ is the capacity of the link between eNBs j and j' , and r_k and r_m represent the required transmission rate in a link for FD call k or group call m .

5) DEFINITION OF VARIABLE r_{bh}

Once we have established the bandwidth constraints, it is easy to define the total use of resources in the backhaul by aggregating the right-hand sides of constraint (16) when multicast is not used in the network or (17) when it is used for all the links in the network. That is, without multicast we have

$$r_{bh} = \sum_{m \in \mathcal{G}} r_m f_{j,j'}^{m,B \rightarrow E} + \sum_{m \in \mathcal{G}} \sum_{i \in \mathcal{U}_m} r_m f_{j,j'}^{m,E \rightarrow B_i} + \sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_k \left(f_{j,j'}^{k,B_i \rightarrow E} + f_{j,j'}^{k,E \rightarrow B_i} \right) \quad (18)$$

while with multicast

$$r_{bh} = \sum_{m \in \mathcal{G}} r_m \left(f_{j,j'}^{m,B \rightarrow E} + f_{j,j'}^{m,E \rightarrow B} \right) + \sum_{k \in \mathcal{T}} \sum_{i=U_k^{(1)}, U_k^{(2)}} r_k \left(f_{j,j'}^{k,B_i \rightarrow E} + f_{j,j'}^{k,E \rightarrow B_i} \right) \quad (19)$$

C. COMPLEXITY ANALYSIS

Theorem 1: The IOPS deployment problem is NP-complete.

Proof: The NP-completeness can be proved by restriction, that is by showing that our general IOPS deployment problem contains a known NP-complete problem as a special case. The reference problem we use in the proof is the multidimensional knapsack problem which is known to be NP-complete. The proof is based on specifying the additional restrictions to be added to the IOPS deployment problem so that the resulting restricted problem will be identical to the multidimensional knapsack problem, which is defined and explained next.

The multidimensional knapsack problem is a well-known problem in combinatorial optimization. It consists of selecting a subset of a given set of items in such a way that the total profit of the selected objects is maximized, while a set of knapsack constraints are satisfied. Formally, the problem can be stated as:

$$\max \sum_{j \in N} p_j x_j \quad (20)$$

$$\text{s.t.} \sum_{j \in N} w_{ij} x_j \leq c_i, \quad i = 1, \dots, d \quad (21)$$

where N is the set of items, d is the number of constraints in the knapsack, p_j is the profit of putting item $j \in N$ in the knapsack, x_j is a binary decision variable indicating whether item j is put into the knapsack, c_i is the capacity of the knapsack in the i -th constraint, and w_{ij} is the weight of item j in the i -th constraint of the knapsack.

To show that our IOPS deployment problem reduces to a multidimensional knapsack problem in some cases, let us consider the following instance of the IOPS deployment problem characterized by the following settings: $\delta = 0$ (we neglect the use of the backhaul in the objective function), $\mathcal{T} =$

\emptyset (we assume we only have group calls), $|\mathcal{U}_m| = 1, \forall m \in \mathcal{G}$ (we assume we have only one UE in each group call). With this, the objective function reduces to

$$\max \sum_{m \in \mathcal{G}} p_m g_m \quad (22)$$

where g_m is a binary variable indicating if group call m is established in the network and p_m is the preference weight (that is, the profit) of group call m , which is of the same form as that in (20).

Let us further assume that: i) each UE is under the coverage of just one eNB, i.e. $|B_i| = 1$; ii) the position of the local EPC is fixed, i.e. h_j is no longer a decision variable, and iii) the path from each eNB to the local EPC is fixed and known. Thus, if a group call m is deployed in the network, it will consume transmission resources in the radio access of the eNB to which the unique UE of group m can be connected, as well as in the links of the path from that eNB to the local EPC. Formally, the IOPS deployment constraints reduces to

$$\sum_{m \in \mathcal{G}} r_{u,j,m} g_m \leq c_{u,j} \quad \forall j \in \mathcal{B} \quad (23)$$

$$\sum_{m \in \mathcal{G}} r_{d,j,m} g_m \leq c_{d,j} \quad \forall j \in \mathcal{B} \quad (24)$$

$$\sum_{m \in \mathcal{G}} r_{j,j',m} g_m \leq c_{j,j'} \quad \forall j \in \mathcal{B}, j' \in \mathcal{E}_j \quad (25)$$

where $r_{u,j,m} = 0$ for all j different from the eNB to which the unique UE of group m can be connected and $r_{u,j,m} = r_{u,m}$ for that eNB. The same applies to $r_{d,j,m}$ for DL. Similarly, $r_{j,j',m} = 0$ for the links (j, j') that are not in the path from the eNB to which the UE of group call m can be connected and $r_{j,j',m} = r_m$ for the rest.

Clearly the previous constraints are similar to those in (21) and the formulation of this particular instance of the IOPS deployment problem matches the previously defined multidimensional knapsack problem, which is known to be NP-complete. By restriction, the IOPS deployment problem must also be NP-complete. \square

D. HEURISTIC ALGORITHM

The IOPS deployment problem is a Mixed Integer Linear Programming (MILP) problem which has been shown to be NP-complete in the previous section. As a consequence, the time required to solve it increases very quickly as the size of the problem grows; therefore, it may not be scalable.

The main objective of solving this problem is to determine the best location for the local EPC. Thus, in this section we propose a heuristic iterative algorithm to estimate that location. The algorithm follows a classical approach based on the Linear Programming (LP) relaxation of the original MILP problem. In short, the algorithm is based on the iterative resolution of simplified relaxed LP problems. In each relaxed problem, we fix $h_j = 1$ for a specific eNB (that is, we force it to be the local EPC of the network), and put the rest of $h_{j'}$ to 0 (line 3 in Algorithm 1). Then, we relax the rest of the binary

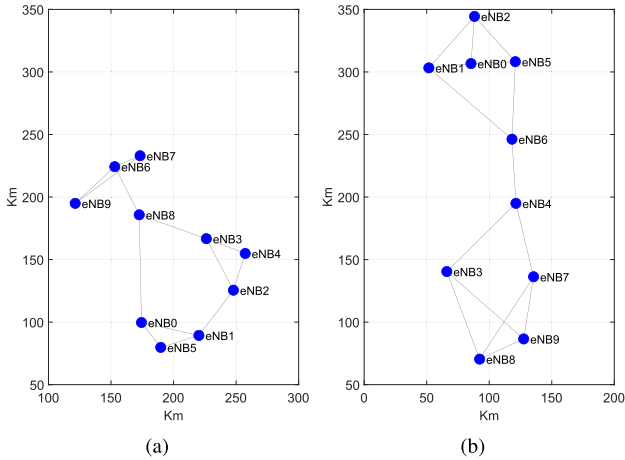


FIGURE 4. Reference deployments. a) Topology A. b) Topology B.

variables to real variables in the range $[0, 1]$ to have an LP problem that can be easily solved (line 4). At each iteration, we check if the objective function of the relaxed problem is the highest so far, updating in that case the candidate eNB to deploy the local EPC (lines 5-8). Thus, we select the eNB that obtains the highest solution of the relaxed problem as the local EPC.

Algorithm 1 Algorithm for Local EPC Placement

input : Instance of the IOPS deployment problem
output: eNB for local EPC placement (EPC_{best})

- 1 $f_{obj, highest} \leftarrow 0$;
- 2 **for** j in \mathcal{B} **do**
- 3 Set $h_j = 1$ and $h_{j'} = 0$ for $j' \neq j$;
- 4 Solve relaxed LP problem;
- 5 **if** $f_{obj, relaxed LP} > f_{obj, highest}$ **then**
- 6 $f_{obj, highest} \leftarrow f_{obj, relaxed LP}$;
- 7 $EPC_{best} \leftarrow j$;
- 8 **end**
- 9 **end**

V. PERFORMANCE EVALUATION

In this section, we evaluate how the proposed model allocates the local EPC in the IOPS mode under different scenarios. The results were obtained by solving the optimization model described in Section 4 using the CPLEX software. The simulated scenarios consider different topologies, traffic of several applications, and transmission schemes (PTT and FD, multicast vs. unicast). In all cases, results were obtained by averaging the outcomes of 100 realizations. Two reference topologies, topology A and topology B, illustrated in Figs 4(a) and 4(b), were considered. The positions of the 10 eNBs in both topologies correspond to the actual deployment of a public safety network managed by an actual PS network operator. In both cases, each eNB establishes up to three backhaul links with neighboring eNBs.

TABLE 2. Transmission rates at backhaul and radio access.

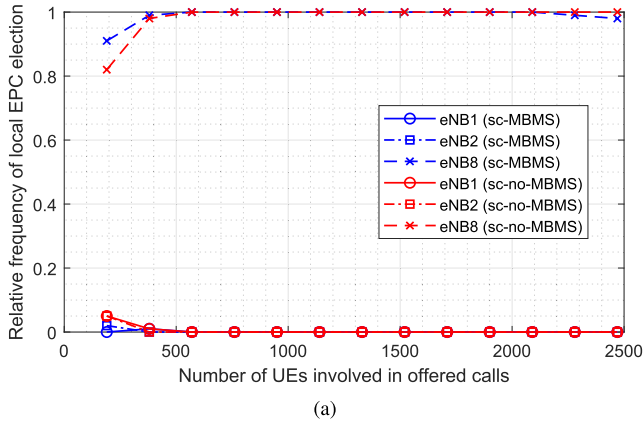
Codec	Backhaul	Radio access
Voice AMR-WB IF1 - 25.2 kbps	41.2 kbps	52 kbps
Video H264 AVC - 350kbps	360 kbps	420 kbps

A 5 MHz LTE deployment was assumed. For simplicity and without loss of generality for the purpose of the local EPC placement problem, we assume an MCS index of 14 for all radio transmissions in both UL and DL (up to 28 MCS index are possible), which leads to an overall transmission rate of 6.36 Mbps in UL and 6.48 Mbps in DL. Excluding signal and control channels (broadcast (BCH), synchronization (PSS/SSS) channels, system information, PDCCH/PUCCH, random access channels (RACH) and signaling overhead (RRC, RSRP reports, etc.), we finally consider an available transmission rate of $c_{u,j} = 5.3$ Mbps in the UL and $c_{d,j} = 5.8$ Mbps in the DL for all the eNBs. Regarding the backhaul links, we consider a transmission rate of $c_{j,j'} = 15$ Mbps in each link, which matches the transmission rate of an LTE link with an MCS index of 26.

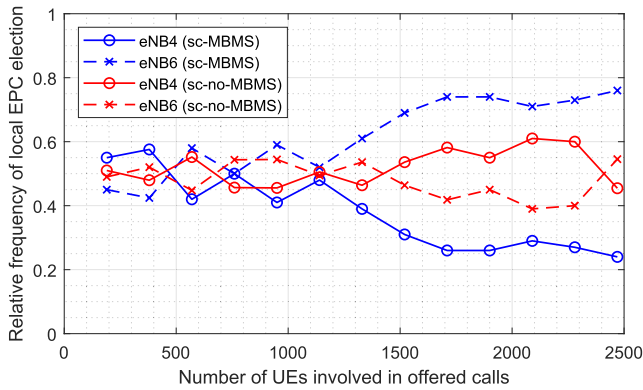
Three different traffic types, representative of PS deployments, are included: PTT video group calls (GC-MCVideo), PTT voice group calls (GC-MCVoice) and FD voice calls (Voice-FD). We evaluate the proposed model to locate the local EPC under two different scenarios: either multicast DL transmissions are considered for group calls (assumed to be sc-MBMS from now on) or unicast transmissions are assumed (denoted as sc-no-MBMS hereafter). Different simulations were performed varying the overall amount of traffic. The ratio of the number of calls for different traffic types was kept fixed at 1:2:4 (GC-MCVideo:GC-MCVoice:Voice-FD). The number of UEs per group is set to 10, also following typical traffic patterns of an actual PS network operator. Transmission rates for both video and voice services at the air interface and backhaul are listed in Table 2. The required backhaul transmission rates include overhead due to transport and network headers (IP, UDP, RTP) while radio access transmission rates, obtained for MCS 14, also include overhead related to radio protocol stack and resource mapping. The preference weights of the different services, p_k and $p_{m,i}$ in (1), are set to 1 and the penalty factor, δ in (1), associated with the use of the backhaul is set to 10^{-3} .

Concerning LTE signaling, as we refer in previous sections, it depends on the particular scenario, user's activity and mobility. Because there are not enough details to accurately model and quantify the signaling and its impact is negligible in front of the data, we only consider it as an additional overhead of data traffic.

UEs are initially attached to one eNB according to a uniform random distribution among all the eNBs of the deployment. Nevertheless, in order to model a certain degree of flexibility we introduce a heuristic to manage a subset of UEs that will be placed in the overlapped coverage areas with a neighbor eNB. UEs in these areas may camp in the initially



(a)

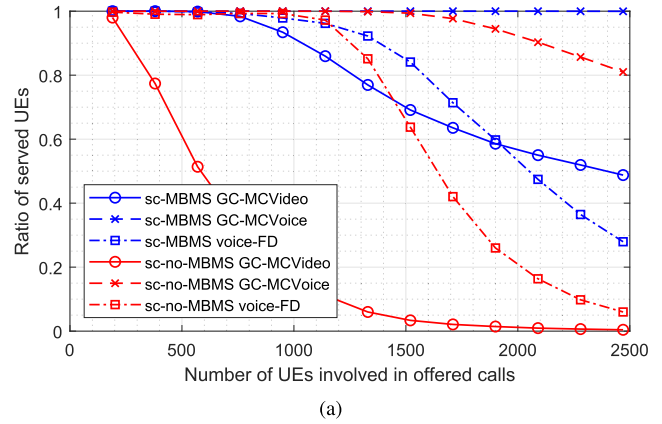


(b)

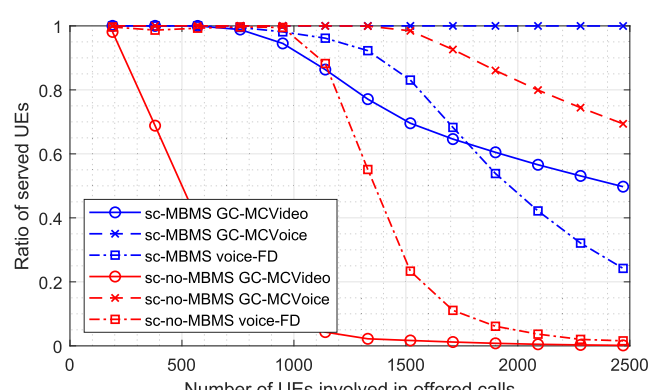
FIGURE 5. Relative frequency of local EPC election under different traffic conditions. (a) Topology A. (b) Topology B.

selected eNB or the neighbor to load balancing by applying an idle mode reselection mechanism [31]. As a result, UEs in these areas are allowed to reselect the eNB where it will finally establish the radio connection. We define the set of possible alternate eNBs (up to three) to the initially selected among i) the closest eNB to the initial one and ii) all eNBs at a distance from the initial eNB less than 10% greater than that of the closest eNB. We assume that a UE attached to the initial eNB can also be within the coverage of one of each of the eNBs in the set with probability $p = 0.1$, and therefore can be reallocated to the latter. That is, if there is a number N of alternate eNB (up to three), a UE may be in the overlapped area of each eNB with the initial one with probability $p = 0.1$ or only under the original cell coverage with probability $1-N.p$.

The traffic is generated and distributed as follows. First, each UE originating a call (group call or FD call) is randomly selected from an eNB according to a uniform distribution. In FD calls, the UE receiving the call is also chosen from a randomly selected eNB. Thus, it is possible to obtain intra-eNB calls and inter-eNB calls. On the other hand, considering the dependence of the location in group calls, the remaining UEs are randomly chosen from the same eNB as the originator or one of the two closest ones. As a consequence, density of UEs involved in group calls at each eNB



(a)



(b)

FIGURE 6. Ratio of served UEs for each type of traffic. (a) Topology A. (b) Topology B.

is not uniform, but depends on the specific topology. For example, in topology B, the probability distribution of GC UEs at eNB i from $i = 0$ to 9 is 0.13; 0.07; 0.1; 0.07; 0.1; 0.13; 0.07; 0.13; 0.1; 0.1.

Fig. 5 shows the probability of locating the local EPC at an eNB for both topologies A (Fig. 5(a)) and B (Fig. 5(b)) under different traffic conditions. These traffic conditions are presented in terms of number of UEs involved in the offered calls. They vary from 190 active UEs (50 involved in 5 GC-MCVideo calls, 100 in 10 GC-MCVoice and 40 in 20 Voice-FD) to 2470 active UEs (650 in 65 GC-MCvideo, 1300 in 130 GC-MCVoice and 520 in 260 Voice-FD) with 100 realizations for each of them. Notice that the portion of UEs involved in MCVideo calls is 5/19, in MCVoice is 10/19 and that in Voice-FD is 4/19. As can be seen, the results in both cases differ owing to the different relative positions among eNBs. The two mixed traffic scenarios, differentiated by GC support, are identified as sc-MBMS and sc-no-MBMS. Fig. 5(a) shows that under this traffic distribution, the local EPC must be unequivocally allocated in eNB8 for both scenarios sc-MBMS and sc-no-MBMS. On the other hand, according to Fig. 5(b), both eNB4 and eNB6 could allocate local EPC for topology B. As can be seen, when multicast is not used, both eNBs are almost equiprobable,

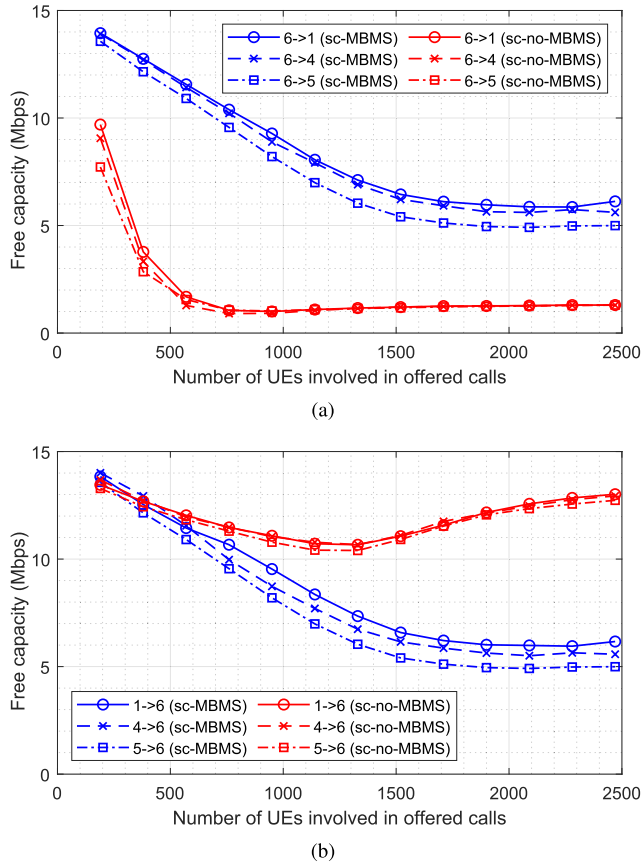


FIGURE 7. Available resources in the backhaul links of eNB8 in topology A. (a) outgoing links. (b) incoming links.

but when multicast is used in the DL, eNB6 tends to be the overall best solution as traffic grows. The reason behind this behavior can be explained as follows: it can be easily seen that topology B is symmetric as long as the backhaul connectivity pattern is concerned. However, as explained above, both the eNB reselection pattern of UEs in the overlapping areas of two cells and the distribution of UEs involved in group calls directly depends on the physical distances among eNBs; therefore, the actual deployment is not exactly symmetric. In this case, when the network is saturated and MBMS is used, the specific overlapping configuration among cells (and thus, cell reselection flexibility) combined with the specific GC distribution probability among cells, results in a slightly higher value in the objective function in the eNBs connected through eNB6 (0, 1, 2 and 5) than in the eNBs connected through eNB4 (3, 7, 8 and 9). There is a slightly higher number of admitted UEs or slightly lower resource utilization in the backhaul due to the higher flexibility provided by the reselection in this area and thus load balancing. Therefore, there is a statistical bias toward eNB6 as a local EPC. Nevertheless, in the actual deployment, both eNBs should allocate the local EPC functionality, although only one would be active.

Fig. 6 shows the fraction of UEs of each traffic type that have been able to establish the connection (denoted as percentage of served UEs), as the mixed traffic increases for both

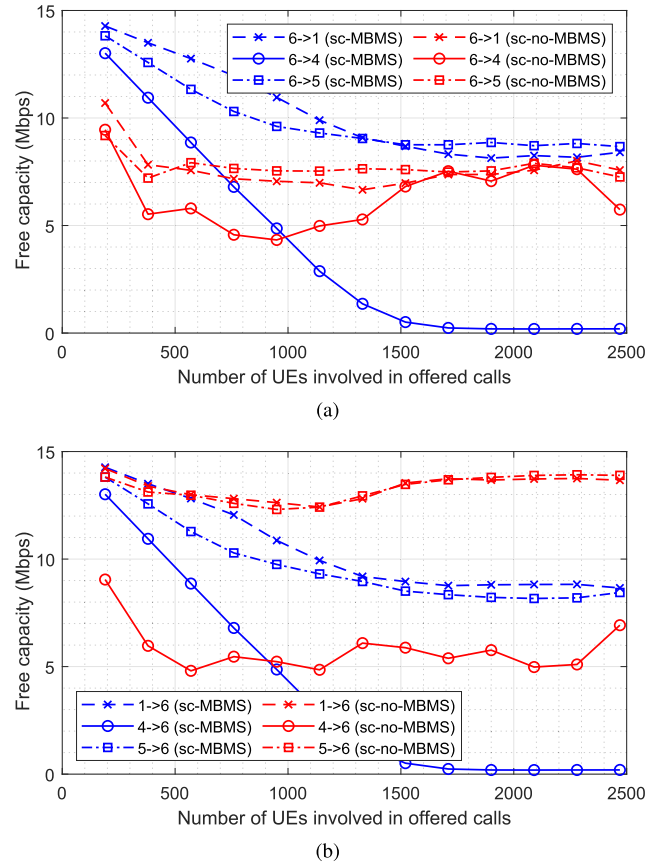


FIGURE 8. Available resources in the backhaul links of eNB8 in topology B. (a) outgoing links. (b) incoming links.

topology A (Fig. 6(a)) and topology B (Fig. 6(b)). It must be noted that whereas Voice-FD calls either are admitted or rejected, group calls can be established even if resources to join the group call for all the involved UEs cannot be allocated. Results show that using multicast in DL transmissions highly improves network capacity due to the reduction in resource consumption both in the radio access and in the backhaul links. The differences in the grade of satisfaction obtained by the different applications lies in the objective function used in the optimization process: the main objective is to maximize the total number of FD calls and the total number of active users in group calls. Because video users consume more resources, they are the first ones to be rejected when the load increases. This is especially obvious when multicast is not used in the DL. Note that in this scenario the decreasing slope for the FD calls is higher than for group calls (also supported using unicast), because of several possible reasons: the main objective function is to maximize the number of active UEs. Considering a number of FD calls that involve a number of UEs equal to a group call (i.e. 5), the number of resources required by the group of FD calls is higher than the equivalent GC. In FD, used resources in the UL are equal to the number of active calls, whereas in the group call only one UE (at least in a cell) is speaking. On the other hand, when the network becomes saturated, the blocking probability for a Voice-FD call can be not only

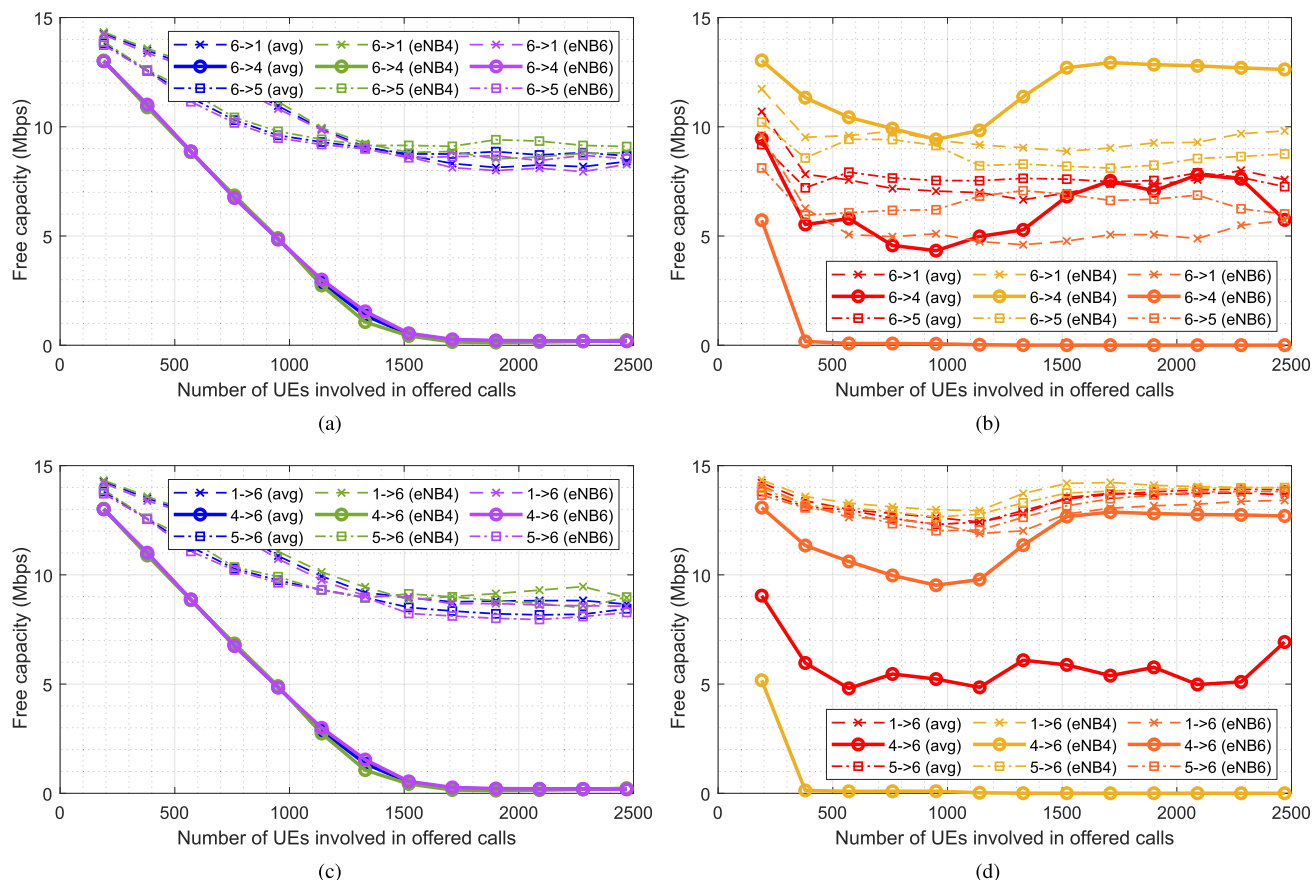


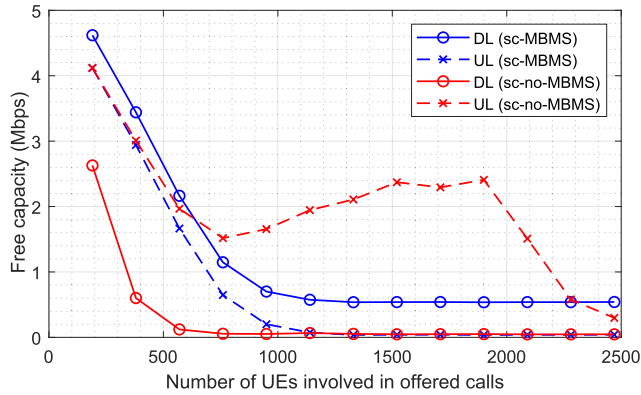
FIGURE 9. Available resources in the backhaul links of eNB8 in topology B decoupled according to the election of local EPC (eNB4 or eNB6). (a) outgoing links sc-MBMS. (b) outgoing links no-sc-MBMS. (c) incoming links sc-MBMS. (d) incoming links no-sc-MBMS.

higher than for a Voice-GC, but also for a Video-GC, even though the resource consumption per UE is lower in voice than video. As stated above, Voice-FD calls cannot be admitted unless there are resources for both the originating and destination UE, whereas a group call can be admitted even if only some (at least two) of the involved UEs can be allocated resources. Thus, GCs are prioritized by the objective function. The resulting network planning with the selected objective function matches the typical requirements of public safety deployments, where GC-MCVoice-PTT calls are the highest priority services in any case.

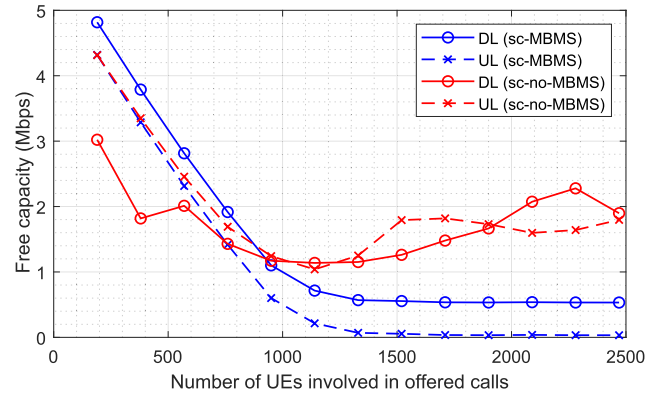
To better understand the system performance and the effects of the local EPC placement, we analyze free resources in backhaul and air interfaces. Fig. 7 represents the amount of free resources in the backhaul links of eNB8, which is the local EPC in topology A. Each backhaul link is represented as transmitter->receiver in the legends, i.e., 6->1 represents the available resources in the transmission link from eNB6 to eNB1. As can be seen in this figure, backhaul links are not the limiting factor in the system capacity since even for high traffic loads there are still free resources. In sc-MBMS, where MBMS is used for GC, resource utilization is similar in both incoming and outgoing links of eNB8. This is because the full duplex consumes the same amount of resources in the

UL and DL by definition. On the other hand, group calls also have a similar utilization: in the UL, only one PTT user per call is active at a time, whereas in the DL, because multicast is used, there is also a single transmission per call in the outgoing links from the local EPC. In sc-no-MBMS, where unicast transport is used for GC, resource utilization is much higher in the DL. Consequently, fewer UEs can be actually admitted (see Fig. 6(a)), and therefore, resource utilization in the incoming links to the local EPC is lower. Consistent with user satisfaction for each class of traffic in Figs 6(a) and 6(b), in Fig. 7(a) and Fig. 7(a), we see that free resources both in incoming and specially in outgoing links increase when FD calls begin to be blocked in favor of GCs. The number of active UEs involved in calls increases with the traffic load, but more UEs of GC are active and they consume less resources.

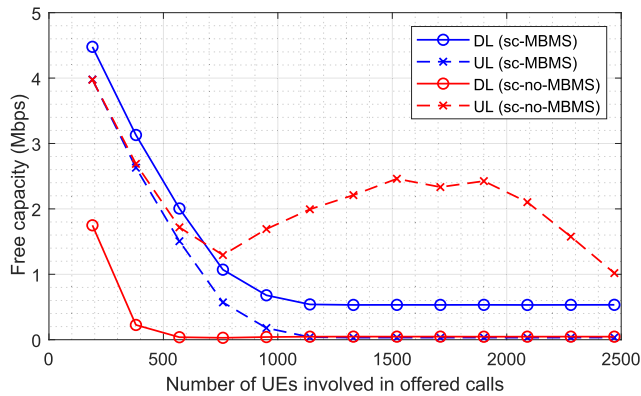
Fig. 8 shows the same results for the backhaul links of the eNB6, one of the two BSs most selected as a local EPC in topology B. It must be noted that the results for eNB4 are equivalent to those of eNB6. Note that results are represented for eNB6, but traffic distributions highly differ depending on the actual eNB co-locating the active local EPC (mainly eNB4 and eNB6), and the results in Fig. 8 show the average amount of free resources in the 100 realizations performed per step. Therefore, to obtain a better understanding of the



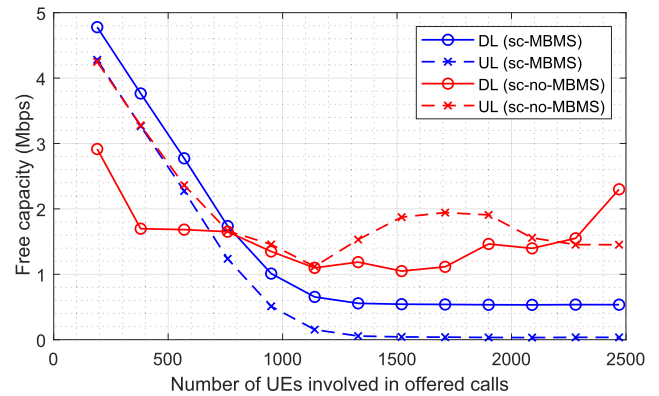
(a)



(a)



(b)



(b)

FIGURE 10. Available resources in the radio access links in topology A. (a) eNB1. (b) eNB6.

actual traffic distribution, we decouple the results in Fig. 9, where we show the same amount of free resources when eNB6 is co-locating the local EPC and when eNB4 is the selected one. In addition, the average values in Fig. 8 are also represented in Fig. 9. For clarity, Figs 9(a) and 9(b) show outgoing links separately for sc-MBMS and sc-no-MBMS, whereas Figs 9(c) and 9(d) show the same results for incoming links. As explained below and unlike topology A, the backhaul link between eNB4 and eNB6 is the limiting factor in topology B.

When MBMS is used (Figs 9(a) and 9(c)), the results are equivalent regardless of the election of local EPC (eNB4 or eNB6), because multicast distribution trees are similar in both cases, as shown in Fig. 4, with a bottleneck link between eNB4 and eNB6. This link is saturated at approximately 1700 UEs. However, when MBMS is not taken into account for GC services (Figs 9(b) and 9(d)), it can be clearly seen that the limiting factor is the DL from the local EPC, that is, the outgoing link from eNB6 to eNB4 when eNB 6 is the local EPC (Fig. 9(b)) or the incoming link to eNB6 from eNB4 when the local EPC is eNB4 (Fig. 9(d)), which is saturated from around 400 UEs. Analogous to topology A, in these cases the corresponding reverse links (4 to 6 and 6 to 4) have more available resources than when the MBMS is used, as fewer calls can be admitted.

FIGURE 11. Available resources in the radio access links in topology B. (a) eNB2. (b) eNB9.

Fig. 10 shows information about resource utilization in the radio access for topology A, namely free radio resources in eNB1 and eNB6. They were chosen as representative examples, because the tendencies are similar for the remaining eNBs. Both eNB1 and eNB6 have a similar density of UEs of group calls, but eNB6 is closer (1 hop) to the local EPC than eNB1 (two hops). Figs 10(a) and 10(b) show the free radio resources in eNB 1 and eNB6 respectively for both UL and DL including sc-MBMS and sc-no-MBMS cases. These results confirm, as stated above in Fig. 7, that the radio access is the limiting factor in this topology: when MBMS is used for GC, the UL is the saturated link, since the assumed maximum transmission rate is slightly lower than in the DL (5.3 Mbps in the UL vs 5.8 Mbps in the DL). If multicast transport is not used for GC, the DL in the radio access is, as expected, the system bottleneck.

Analogously, Fig. 11 shows information about resource utilization at the air interface for topology B. In this case, the results are shown for eNB2 and eNB9. These eNBs have been chosen because of their symmetry regarding the election of the local EPC: they are both remote eNBs that are connected to the remaining deployment either through eNB6 or eNB4 and both eNBs have a similar density of UEs of group calls, as shown in the distribution probability described above.

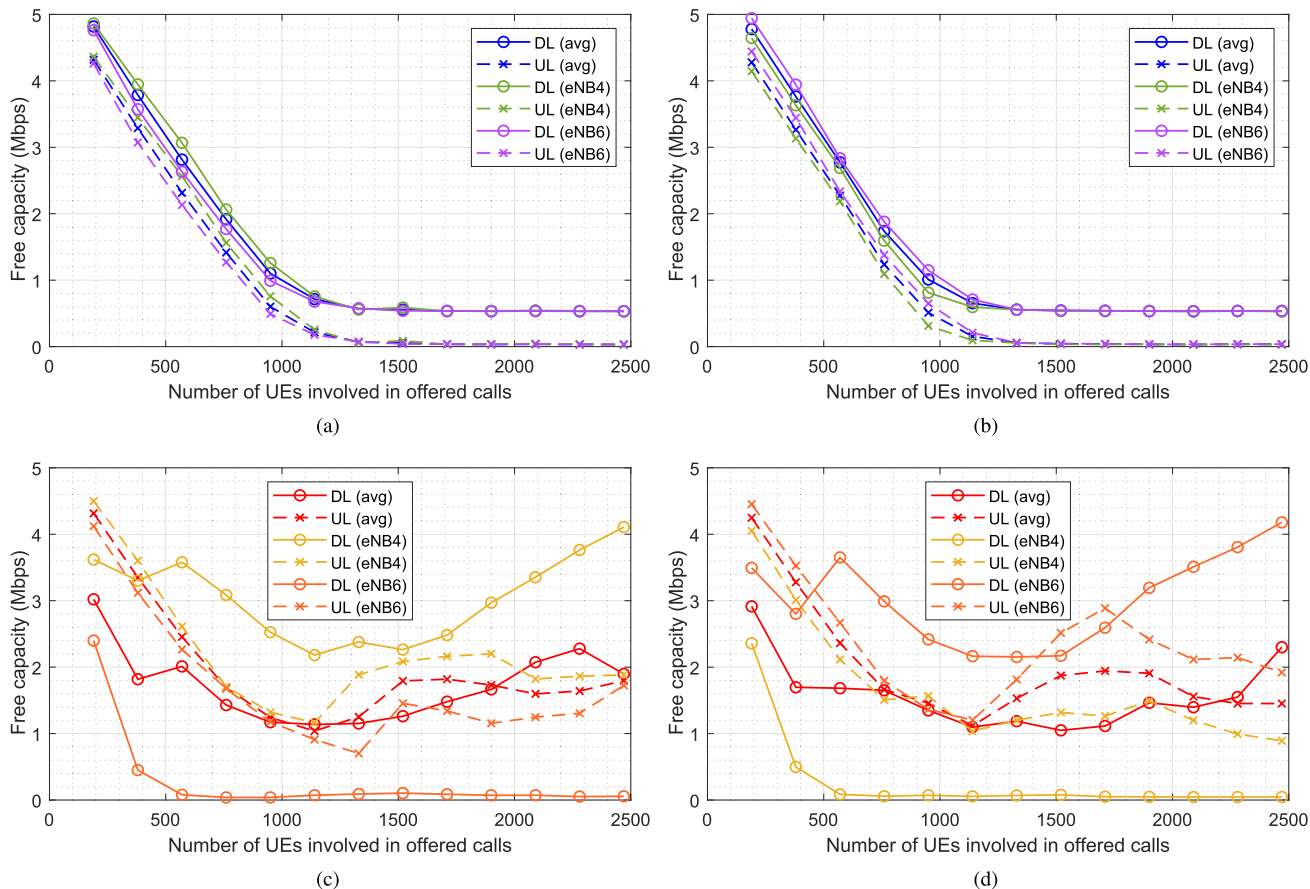


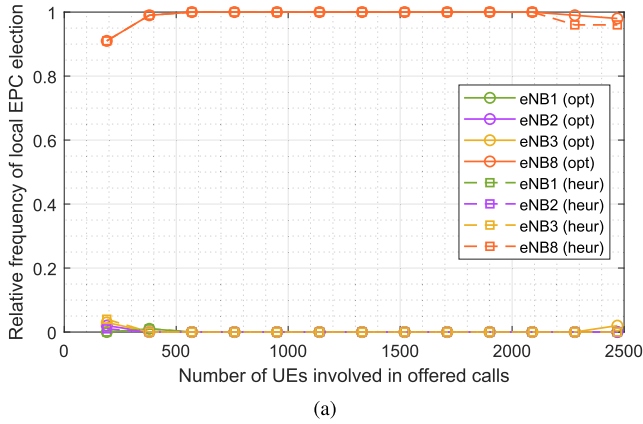
FIGURE 12. Available resources in the radio access links in topology B decoupled according to the election of local EPC (eNB4 or eNB6). (a) eNB2 sc-MBMS. (b) eNB9 sc-MBMS. (c) eNB2 sc-no-MBMS. (d) eNB9 no-sc-MBMS.

Fig. 11(a) shows the available radio resources in eNB2, which is located in the region connected through eNB6, whereas Fig. 11(b) shows equivalent results for eNB9, which is located in the region connected through eNB4. Following the same reasoning as in Fig. 8, we decouple the results in Fig. 12, where we show the free radio resources when eNB6 is co-located with the local EPC and when eNB4 is the selected one, together with the average values of Fig. 11. Again, for clarity, Figs 12(a) and 12(b) show the radio resources for eNB2 and eNB9 for sc-MBMS, whereas Figs 12(c) and 12(d) show the same results sc-no-MBMS.

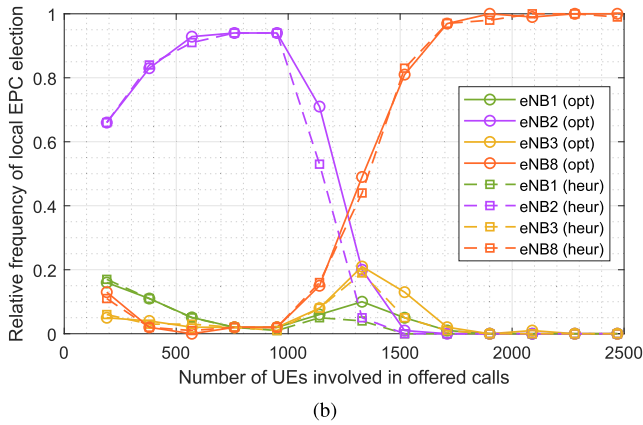
As explained in Figs 9(a) and 9(c) in the backhaul, when MBMS is used, the results are almost similar regardless of the election of local EPC (eNB4 or eNB6), as can be seen in Figs 11(a) and 11(b). Nevertheless, when multicast is not used, the results differ depending on the selected local EPC. If eNB6 is co-located with the local EPC, the performance is limited by the outgoing backhaul link from eNB6 to eNB4 (see Fig. 9(b)). As a result, the blocking probability of UEs that reaches eNB6 through eNB4 is high even if there are enough resources in their air interface. Therefore, in Fig. 12(d), eNBs that are connected to the local EPC through eNB4, such as eNB9, have available resources in

the radio access. The same does not occur in eNBs that are not connected to the local EPC through eNB4, like eNB2. In this case, the bottleneck is not the backhaul, but rather the radio interface. Specifically, the DL radio access is saturated, as shown in Fig. 12(d) for eNB2. Likewise, when eNB4 is co-located with the local EPC, the bottleneck is located in the outgoing backhaul link from eNB4 to eNB6 (see Fig. 9(d)). Thus, in Fig. 12(b), eNBs that are connected to the local EPC through eNB6, such as eNB2, have available resources in the radio access although no more calls can be supported, while in eNBs not connected to the local EPC through eNB6, the DL radio access is saturated, as shown in Fig. 12(d) for eNB9.

As shown in Fig. 5, from the point of view of the optimization function that maximizes the number of satisfied UEs in the overall deployment, the preferred local EPC placement location is equally distributed in eNB4 and eNB6, although as shown above, the selection of eNB4 or eNB6 in sc-MBMS provides differences in the area connected to eNB4 and eNB6. In the sc-no-MBMS scenario, these differences were significantly reduced. Despite this, as shown in Fig. 5, eNB6 is selected as the optimum with a higher probability than eNB4 as the offered traffic increases. This preference, which is difficult to derive from the occupation results, is due to the

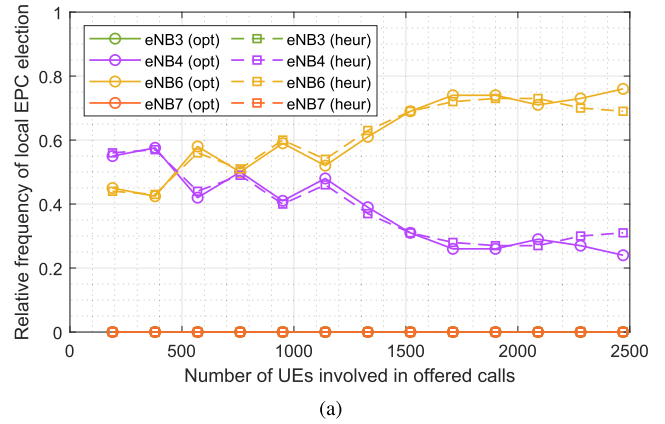


(a)

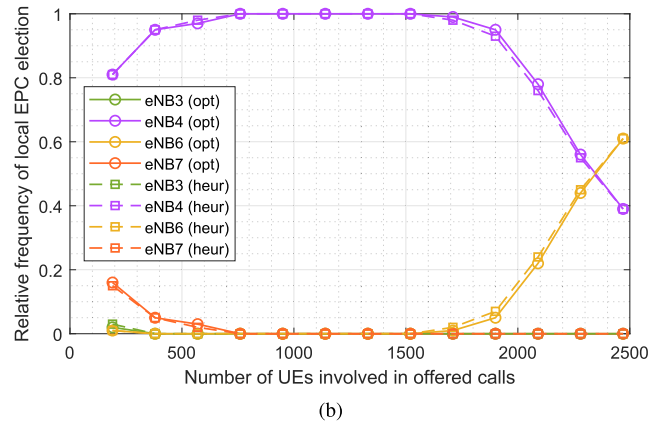


(b)

FIGURE 13. Relative frequency of local EPC election under different traffic conditions in topology A for both optimum and heuristic solutions. (a) initial nearly homogeneous traffic. (b) heterogeneous traffic.



(a)



(b)

FIGURE 14. Relative frequency of local EPC election under different traffic conditions in topology B for both optimum and heuristic solutions. (a) initial nearly homogeneous traffic. (b) heterogeneous traffic.

small differences between the capabilities of cell reselection in the areas connected to eNB4 and eNB6. In this specific deployment, the area connected to eNB6 can balance the traffic among cells slightly better than the area connected to eNB4. Because the connections in sc-MBMB consume much less resources than in sc-no-MBMS, small resource savings may be sufficient to allocate a new UE. Thus, although the number of allocated UEs only differs in some units more, eNB6 tends to be the preferred option.

To better show the dependence of the local EPC allocation, not only on the specific topology, but also on the traffic distribution, we also evaluate the proposed model under more heterogeneous traffic patterns, where each UE originating a call (GC or FD call) is selected from a randomly chosen eNB according to a non-uniform distribution. Specifically, for topology A, the probability distribution of UEs that generate a call at eNB i from $i = 0$ to 9 is 0.05; 0.15; 0.15; 0.15; 0.15; 0.15; 0.05; 0.05; 0.05; 0.05, while for topology B, the distribution is 0.07; 0.07; 0.07; 0.13; 0.13; 0.07; 0.07; 0.13; 0.13; 0.13. These probability distributions differ from the probability distributions of the UEs involved in GC. Once the distribution of the UEs that generate the call is established, they depend on the heuristic defined above.

For clarity and without loss of generality in the conclusions, we now focus on the sc-MBMS scenario. Figs 13 and 14 show the probability of locating the local EPC at an eNB for both topologies under the initial nearly homogeneous traffic deployment (a) and the aforementioned heterogeneous patterns (b). Joint to the optimal solution, we include the solution provided by the heuristic algorithm proposed in Section IV-D. It can be seen that the results obtained by the heuristic algorithm are close to the optimal one, with a much lower computational complexity. Therefore, it can be a useful tool when network topology grows to obtain suboptimal solutions with limited computation time. Regarding the nearly homogeneous traffic deployment, Figs 13(a) and 14(a) show the results already shown in Fig. 5 compared with the results obtained by the heuristic algorithm. As for the heterogeneous traffic pattern concerns in topology A, the traffic density in eNBs 1, 2, 3, 4 and 5 is higher than in the remaining ones. Thus, we can see in Fig. 13(b) that the local EPC placement initially moves toward the center of the area where these five eNBs are located; thus, the optimal placement location is at eNB2 in most cases. However, as traffic grows and the network becomes saturated in all eNBs, and not only in those with higher density, the local EPC moves again to eNB8, because this location allows for an overall

higher number of calls. Similarly, the heterogeneous traffic pattern in topology B increases the traffic density in the area of eNBs connected through eNB4 (3, 4, 7, 8, and 9). Thus, the local EPC is located in eNB4 (Fig. 14(a)) instead of the duality between eNB4 and eNB6 seen in the homogeneous scenario (Fig. 14(a)). However, with the same argument as in topology A, as the network becomes saturated, the local EPC moves again to eNB6, which is also the most selected location as traffic grows in Fig. 14(b).

Obviously, the obtained results depend on the topology, traffic mix, and UEs distribution, in addition to the specific mode of provision for GC on the deployment in IOPS operation mode. Both the optimization and the heuristic algorithms need to be reconfigured using feedback with input data of the traffic mix and the distribution in the area of provision, obtained or estimated from previous operative scenarios. In any case, from the analysis performed above we can conclude that, probably in many topologies, planning standby IOPS co-located in all eNBs to respond to various traffic scenarios when a group of eNBs enter the IOPS mode will not be needed, but rather in a set of them. Specifically, in the topologies analyzed above, some eNBs should always be part of this set. This is the case for eNB8 and eNB6 in topologies A and B, respectively, because in saturation conditions these placements maximize the number of UEs participating in calls. In addition, there are options with almost equivalent benefits that should also be included in the set. This is the case of eNB4 in topology B. In addition, other eNBs such as eNB2 or even eNB1 and eNB3 may be included to respond to traffic variations in topology A.

Then, the choice of active/preferred local EPC must be based on local policies that consider both the previous load conditions that the area suffers when entering the IOPS mode, and expected increases based on previous experiences. For instance, eNB2 could be set as the preferred option in topology A in the heterogeneous scenarios defined above when the load is moderate. The current optimization function maximizes the UEs on call without considering whether only two or all members of the GC are satisfied. We are aware that the optimization function can be changed to ensure that a minimum number of UEs in the GC have resources for the call. In this case, the required changes in the optimization algorithm were minimal.

VI. CONCLUSION

Future PS networks will use commercial standards such as 4G/5G to provide broadband services. Therefore, the 3GPP has introduced specific amendments to support the services needed in such networks, such as proximity communications, PTT group communications, or the possibility of operating in IOPS mode. In this paper, we proposed an optimization problem to find the best location to co-locate the local EPC within a set of eNBs working in the IOPS mode. This optimization problem models specific features of PS networks, such as PTT and, specifically, group call communications, using different delivery modes (unicast and MBMS), as well as com-

mon voice PtP communications. Owing to the NP-hardness of the problem, a low-complexity heuristic algorithm has been proposed to estimate the optimal location of the local EPC. Simulation results show that the local EPC tends to be co-located with one of the geographically central eNBs of the network. Nevertheless, when the traffic is heterogeneous, the best location for the local EPC depends on the specific spatial distribution of UEs in different communications. Finally, our heuristic algorithm shows that the local EPC placement problem can be satisfactorily solved with a low computational burden.

REFERENCES

- [1] (2013). *Public Safety on 3GPP*. Accessed: May 19, 2021. [Online]. Available: <https://www.3gpp.org/news-events/1455-public-safety>
- [2] *Study on Architecture Enhancements to Support Group Communication System Enablers for LTE (GCSE_LTE) (Rel. 12)*, Standard 3GPP TR 23.768 V12.1.0, Jun. 2014.
- [3] *Proximity-Based Services (ProSe); Stage 2, (Rel. 12)*, Standard 3GPP TS 23.303 v.12.8.0, Mar. 2015.
- [4] *Mission Critical Push To Talk (MCPTT) Over LTE; Stage 1 (Rel. 13)*, Standard 3GPP TS 22.179 V13.3.0, Dec. 2015.
- [5] (2017). *Mission Critical Services in 3GPP*. Accessed: May 19, 2021. [Online]. Available: https://www.3gpp.org/news-events/1875-mc_services
- [6] *Study on Architecture Enhancements to Support Isolated Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Operation for Public Safety (Rel. 13)*, Standard 3GPP TR 23.797 V13.3.0, Dec. 2015.
- [7] S.-Y. Lien, C.-C. Chien, G. S.-T. Liu, H.-L. Tsai, R. Li, and Y. J. Wang, "Enhanced LTE device-to-device proximity services," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 174–182, Dec. 2016.
- [8] J. Oueis, V. Conan, D. Lavaux, R. Stanica, and F. Valois, "Overview of LTE isolated E-UTRAN operation for public safety," *IEEE Commun. Standards Mag.*, vol. 1, no. 2, pp. 98–105, Jul. 2017.
- [9] *Functional Architecture and Information Flows to Support Mission Critical Video (MCVideo); Stage 2*, Standard 3GPP TS 23.281, 2016.
- [10] *Functional Architecture and Information Flows to Support Mission Critical Data (MCData); Stage 2*, Standard 3GPP TS 23.282, 2017.
- [11] *Mission Critical Services Support in the Isolated Operation for Public Safety (IOPS) Mode of Operation; Functional Architecture and Information Flows; (Rel. 17)*, Standard 3GPP TS 23.180 V17.0.0, Sep. 2020.
- [12] *Group Communication System Enablers for LTE (GCSE LTE); Stage 2*, Standard 3GPP TS 23.468, 2014.
- [13] *Common Functional Architecture to Support Mission Critical Services. Stage 2*, Standard 3GPP TS 23.280, 2016.
- [14] J. Kim, S. W. Choi, W.-Y. Shin, Y.-S. Song, and Y.-K. Kim, "Group communication over LTE: A radio access perspective," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 16–23, Apr. 2016.
- [15] *Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description*, Standard 3GPP TS 23.246, 2007.
- [16] *Study on Single-Cell Point-to-Multipoint Transmission for E-UTRA*, Standard 3GPP TR 36.890, 2015.
- [17] S. Feng, C. Liu, C. Shen, H.-A. Choi, and R. A. Rouil, "An effective and efficient dynamic eMBMS multicast grouping scheduling algorithm in MBSFNs for public safety scenarios," *IEEE Access*, vol. 8, pp. 105701–105712, 2020.
- [18] D. Li, P. Hong, K. Xue, and J. Pei, "Virtual network function placement and resource optimization in NFV and edge computing enabled networks," *Comput. Netw.*, vol. 152, pp. 12–24, Apr. 2019.
- [19] D. Bhamare, M. Samaka, A. Erbad, R. Jain, L. Gupta, and H. A. Chan, "Optimal virtual network function placement in multi-cloud service function chaining architecture," *Comput. Commun.*, vol. 102, pp. 1–16, Apr. 2017.
- [20] A. Mouaci, R. Gourdin, I. Ljubic, and N. Perrot, "Virtual network functions placement and routing problem: Path formulation," in *Proc. IFIP Netw. Conf. Netw.*, Jun. 2020, pp. 55–63.
- [21] J. Oueis, R. Stanica, F. Valois, V. Conan, and D. Lavaux, "Core network function placement in mobile networks," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

- [22] J. Oueis, V. Conan, D. Lavaux, H. Rivano, R. Stanica, and F. Valois, "Core network function placement in self-deployable mobile networks," *Comput. Commun.*, vol. 133, pp. 12–23, Jan. 2019.
- [23] J. Oueis, R. Stanica, and F. Valois, "Virtualized local core network functions placement in mobile networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
- [24] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Netw.*, vol. 1, no. 3, pp. 215–239, 1979.
- [25] A. Bavelas, "Communication patterns in task oriented groups," *J. Acoust. Soc. Amer.*, vol. 22, no. 6, pp. 271–282, 11 1950.
- [26] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977.
- [27] F. Li, Y. Wang, X.-Y. Li, A. Nusairat, and Y. Wu, "Gateway placement for throughput optimization in wireless mesh networks," *Mobile Netw. Appl.*, vol. 13, nos. 1–2, pp. 198–211, Apr. 2008.
- [28] Z. Zheng, L. X. Cai, M. Dong, X. Shen, and H. V. Poor, "Constrained energy-aware AP placement with rate adaptation in WLAN mesh networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.
- [29] Y. Drabu and H. Peyravi, "Gateway placement with QoS constraints in wireless mesh networks," in *Proc. 7th Int. Conf. Netw. (ICN)*, Apr. 2008, pp. 46–51.
- [30] B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2127–2136, Nov. 2006.
- [31] S. Mishra and N. Mathur, "Load balancing optimization in LTE/LTE—A cellular networks: A review," Dec. 2014, *arXiv:1412.7273*.



JORGE ORTÍN was born in Zaragoza, Spain, in 1981. He received the M.S. and Ph.D. degrees in engineering of telecommunications from the Universidad de Zaragoza, in 2005 and 2011, respectively.

In 2008, he joined the Aragón Institute of Engineering Research, Universidad de Zaragoza, where he has participated in different projects funded by public administrations and by major industrial and mobile companies. In 2012, he joined the Universidad Carlos III of Madrid, Madrid, Spain, as a Postdoctoral Research Fellow. Since 2013, he has been working at the Centro Universitario de la Defensa Zaragoza. His research interest includes wireless communications systems.



JOSÉ RAMÓN GÁLLEGO was born in Zaragoza, Spain, in 1978. He received the M.S. and Ph.D. degrees in engineering of telecommunications from the Universidad de Zaragoza, Zaragoza, in 2001 and 2007, respectively.

In 2002, he joined the Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, where he is currently an Associate Professor and a member of the Aragón Institute of Engineering Research. His current research interests include wireless communications, with an emphasis on radio resource management and wireless sensor networks.



ÁNGELA HERNÁNDEZ-SOLANA received the M.S. and Ph.D. degrees in engineering of telecommunications from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1997 and 2005, respectively.

She was with UPC and the University of Zaragoza, Zaragoza, Spain, where she has been an Associate Professor, since 2010. She is currently a member of the Aragón Institute of Engineering Research, UPC. Her research interests include 5G/4G technologies, heterogeneous communication networks and mission-critical communication networks, with emphasis on transmission techniques, radio resource management and quality of service, mobility management and planning, and dimensioning of mobile networks.



MARÍA CANALES was born in Zaragoza, Spain, in 1978. She received the M.S. and Ph.D. degrees in engineering of telecommunications from the Universidad de Zaragoza, Zaragoza, in 2001 and 2007, respectively.

In 2002, she joined the Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, where she is currently an Associate Professor and a member of the Aragón Institute of Engineering Research (I3A). Her current research interests include wireless communications with an emphasis on radio resource management and wireless sensor networks.

• • •