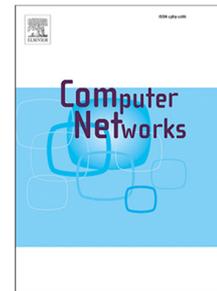


Journal Pre-proof

Power efficient mobile small cell placement for network-coded cooperation in UDNs

Roberto Torre, Muhammad Tayyab, George Koudouridis,
Xavier Gelabert, Riccardo Bassoli, Frank H.P. Fitzek



PII: S1389-1286(21)00451-5
DOI: <https://doi.org/10.1016/j.comnet.2021.108559>
Reference: COMPNW 108559

To appear in: *Computer Networks*

Received date : 12 February 2021
Revised date : 22 July 2021
Accepted date : 1 October 2021

Please cite this article as: R. Torre, M. Tayyab, G. Koudouridis et al., Power efficient mobile small cell placement for network-coded cooperation in UDNs, *Computer Networks* (2021), doi: <https://doi.org/10.1016/j.comnet.2021.108559>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Elsevier B.V. All rights reserved.

Power Efficient Mobile Small Cell Placement for Network-coded Cooperation in UDNs

Roberto Torre^a, Muhammad Tayyab^b, George Koudouridis^c,
Xavier Gelabert^c, Riccardo Bassoli^a, Frank H.P. Fitzek^a

^a*Deutsche Telekom Chair of Communication Networks and Centre for Tactile Internet (CeTi) with Human-in-the-Loop, Georg-Schumann-Strae 11, Dresden, 01187, Saxony, Germany*

^b*Huawei Technologies Finland and Department of Communications and Networking, School of Electrical Engineering, Aalto University, Espoo, Finland*

^c*Huawei Technologies Sweden AB, Stockholm, Sweden*

Abstract

Ultra-dense Networks (UDNs) massively populate areas with base stations of diverse capabilities, thus increasing the network capacity. Moreover, the radio access network (RAN) architecture moves towards small infrastructure elements such as mobile small cells (MSCs). In this context, Network-coded Cooperation (NCC) leverages the interplay between network coding and cooperative relaying to reliably offload cellular traffic to MSCs and reduce the power consumption in the network. Despite the research done separately on NCC and smart MSC deployment, there is a scarceness of works addressing the smart and on-demand deployment, activation, and deactivation of MSCs to leverage the benefits of both NCC and MSCs. To fill this gap, in this paper, we: (1) estimate the traffic density of New York by adopting an urban zoning (UZ) model; (2) provide a methodology for the on-demand deployment of base stations and MSCs according to a stochastic geometry model; (3) propose two radio resource management (RRM) models, one random and one smart, for the placement and on-demand creation of MSCs, and (4) compare the power consumption of the proposed architecture with 4G edge computing. The results show that the smart RRM model overperforms the random model five-fold in terms of number of pico base stations required,

Email address: roberto.torre@tu-dresden.de (Frank H.P. Fitzek)

URL: <https://cn.ifn.et.tu-dresden.de/> (Frank H.P. Fitzek)

which impact on the power consumed in the network. Moreover, the results show that the smart model achieves between 6 – 25% power savings in comparison to 4G edge computing, the random model, and two approaches from the related work, respectively.

Keywords: Network-coded cooperation, mobile small cells, power consumption, cellular networks, stochastic geometry, network traffic, urban zoning

1. Introduction

As a new generation unfolds, the architecture of the network undergoes a series of structural changes from the core network to the user. New infrastructure elements, such as femto/pico Base stations (BSs), fixed/mobile relays, cognitive radios, and distributed antennas are being massively deployed, thus making the future cellular networks beyond 5G more heterogeneous [1]. In this emerging environment, new services and handheld devices increase the number of internet connections and mobile traffic.

Specifically, it is estimated [2] that by 2023, the number of connections will be 29.3 billion and that the number of connected devices will reach 13.1 billion in 2022, 71% of which being mobile devices. Reports also show the steep yearly mobile data traffic growth trend, e.g. 63% from 2015 to 2016 [3], which is expected to continue in years to come. Moreover, a mobile data traffic increase up to 77.5 exabytes per month was reported in 2017 [2], from which 82% of that traffic consisted of IP video traffic. In view of this, and especially in dense urban areas where a massive amount of users coexist, we can argue that today's network infrastructure is not ready for such changes, and that a paradigm shift is needed. In this paper, we tackle the aforesaid problems by jointly considering Network-coded Cooperation (NCC) schemes along with the deployment of Ultra-Dense Networks (UDNs), briefly introduced hereafter.

To cope with the increasing demand of video streaming services, NCC [4] presents an energy efficient method for massive downlink content distribution in the Radio Access Network (RAN). As the name suggests, NCC involves some degree of cooperation (C) between the participating nodes in a communication transmission along with Network Coding (NC) capabilities of such nodes, hence NCC. In this paper, we are concerned with NCC at the RAN,

where arrangements of BSs and mobile devices (i.e. User Equipments (UEs)) cooperate with each other in order to achieve efficient and reliable communication [5]. We will refer to this arrangements as Mobile Small Cells (MSCs) [6] noting that, not only the UEs, but also the BSs are mobile (e.g. via a drone-mounted BS). In this scenario, the challenges reside in the mobility nature of the MSCs, meaning that the arrangement and composition of BSs and UEs in a MSC is dynamic and thus can change over time. Noteworthy, despite the research done in the area of NCC, there is still a lack of algorithms to efficiently deal with these dynamic scenarios. This paper aims to contribute to this direction.

In addition to NCC, an effective way to increase the capacity of cellular systems under high traffic and UE density relies on the large deployment of low-power BSs with short inter-site-distance (ISD). Typically, when the ISD falls below 100m, thus creating small cell coverage regions, we refer to these deployments as UDNs[7, 8]. An additional benefit from UDNs is that network power consumption can be reduced by conveniently switching on and off those BSs with no active UE transmissions. In these scenarios, the majority of adopted traffic models in the literature assume that the offered traffic is constant. However, in reality traffic variations occur in the time domain (e.g. morning/afternoon/evening or weekday/weekend) [9], and in the spatial domain (e.g. residential/commercial/industrial) [10]. Furthermore, traffic models in existing works are largely based on statistics from huge datacenters where all traffic is aggregated. This approach may be good if the scope of the evaluation covers large areas, but it is suboptimal if we look into the traffic variation at each BS separately. As an example, population fluctuations (e.g., due to commuting) can abruptly change the traffic load within a city, and this directly impacts on the service load that is requested from particular BSs. We aim to solve this problem by means of a comprehensive, exhaustive, and more accurate model of the urban area by dividing it into zones. We call this model Urban Zoning (UZ) and it divides the city into three different zones, namely: residential, commercial, and manufacturing. Each zone has already some existing infrastructure with a given number of static micro base stations (mBSs), i.e. BSs with ISDs in the range of hundreds of meters.

As explained above, the current network infrastructure is not prepared for the increase in network traffic from upcoming applications and services. This traffic increase would require the deployment of additional RAN equipment. Unfortunately, changing the number of fixed mBSs would suppose a huge economic and time investment. As an alternative, in this work, the

network traffic surplus is offloaded to MSCs, which is defined as a group of connected mobile devices such as vehicles, drones, high-altitude balloons, or UEs. Within a MSC, one mobile device assumes the role of a pico base station (pBS) which provides small cell coverage and connectivity to the rest of devices conforming the MSC. Among the many possible implementations of a MSC [6], in this work we will assume that an Unmanned Aerial Vehicle (UAV), i.e. a drone, is equipped with a pBS providing small cell coverage to a group of UEs.

Our contribution is four-fold. First, we propose a novel UZ model for the analysis of the network traffic in the RAN; second, we optimize the mBS and pBS density to reduce the number of NCC MSCs to minimize the energy consumption in the network; third, we provide two Radio Resource Management (RRM) heuristics (namely the random model and the smart model) that deal with the placement of pBSs in the given scenarios (i.e. residential, commercial, and manufacturing); fourth, we evaluate the sensitivity of the scenario against changes in user density, pBS range (or pBS transmission power), user mobility, and maximum pBS density and compare the cases with and without MSC deployment. The metrics selected for the sensitivity evaluation are: power consumed in the RAN, service efficiency, user distribution, pBSs user density, pBS density, and pBS coverage.

The results present a low Minimum Squared Error (MSE) between the model and the simulator. Furthermore, the numerical results show total power savings in the network, especially in the RAN, in comparison to the 4G edge computing. Moreover, the smart model is nine to eleven-fold more efficient than the random model, depending on the hour of the day and the urban zone. Finally, we perform a sensitivity benchmark of the smart model (since it is the best one) under the aforementioned variations in the proposed scenario. The sensitivity results show a linear impact of the user expectation and maximum pBS density in the metrics studied. Moreover, the pBS range and mobility of users present different maximums and minimum that show different trade-offs.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the evaluated scenario. Section 4 models the traffic density in a metropoli. Section 5 introduces the power consumption models used in the different parts of the network. Section 6 gives an overview of the restrictions applied to the model in the fronthaul and core network. Section 7 gives an overview of the RAN model. Section 8 proposes an optimization problem to find the optimal number of mBS and

pBS in the network and introduces two RRM heuristics for the placement of pBS in the RAN. Finally, section 9 collects the results of the evaluation.

2. Related Work

2.1. Network-coded Cooperation

Ahlswede et al. [11] introduced NC as a technology that increases throughput and resilience in wireless networks by having intermediate network nodes perform operations rather than mere forwarding of packets. Among the possible NC implementations, Random Linear Network Coding (RLNC) [12] is a popular NC protocol used by streaming services since the intermediate nodes do not need to know the coding coefficients to forward the packets, but rather they can generate new coding coefficients as soon as the packets arrive, which helps reducing the latency of the system. Furthermore, network cooperation [13] has been introduced as a technology that increases performance in wireless networks. Noteworthy, it has been proven that the combination between NC and cooperation increases network performance in terms of throughput and resilience and, as a result, it reduces the energy consumption of the system [14]. In this context, NCC is introduced as a combination of RLNC and cooperative relaying that improves the performance of the network by increasing its throughput and reducing the power consumption. Leiva-Mayorga et al. [15, 16] proposed a Markov Decision Process (MDP) analysis to model the behavior of NCC. Torre et al. [17] corroborated the results proposed by Leiva-Mayorga by means of a testbed, which was further extended into a demonstrator, presented in [5]. In view of the above, we consider NCC as a suitable protocol to enhance the network capacities and support the increase of traffic.

2.2. Ultra-dense Networks and UAVs

Heterogeneous UDNs [7, 8] has become popular in the recent years to deal with the increase of the number of devices and cellular traffic. For example, López-Pérez et al. [18] evaluated the gains in network throughput, energy efficiency, and Signal to Interference and Noise Ratio (SINR) of high-frequency bands due to network densification. On the other hand, Liu et al. [19] addressed UDNs under different scopes such as Heterogeneous networks (HetNets), massive Multiple-input Multiple-output (MIMO), and mmWave networks. Similarly, Lu et al. [20] addressed the topic of UDNs from the perspective of energy harvesting.

There are different forms to deploy a UDN. Predominantly, the majority of existing literature considers the static deployment of base stations, e.g. [21]. However, recent technology advancements enable the placement of static BS functionalities onto mobile devices, thus enabling the MSC paradigm via mobile BSs. Examples of those mobile devices range from medium-sized computing devices on-board vehicles (cars, buses, etc.) to small-sized computing devices like UAVs (drones), Google balloons, or even UEs with enhanced capabilities [22]. For small-sized computing devices, the necessary hardware (i.e. antennas, radio units, etc.) attached to the mobile BSs need to be power-efficient, compact, and lightweight so that mobile BSs can carry and power them. Bearing this in mind, pico and femto base stations become the best candidates to be mounted on mobile devices due to their small size and low power consumption. In this context, Li et al. [23] proposed a context-aware optimization for the location of drone base stations. Huang et al. [24], on the other hand, formulated drone-deployment problems as NP-hard problems and proposed greedy algorithms with their theoretical analysis. Authors therein also considered the drone's battery constraints and the distance on the air. However, they only considered the deployment of drones along the streets. On the other hand, in their work of Li et al. [23], the drones must not necessarily be along the streets. Finally, Kalantari et al. [25] proposed a heuristic method to place drones that act as base stations in an area with different user densities, which attended to the traffic variations over the network. However, the authors did not consider the possibility to move the drones between the areas, attending to dynamic variations of users densities.

The works presented above, except the latter one, modeled the traffic in urban areas with similar user densities. However, it is worth noting that the user density of the network can be also modeled using the server workload. For example, Bassoli et al. [9] used the server traffic to characterize the traffic in the network and to provide a comprehensive comparison between 4G and 5G in terms of power consumption. Similarly, Lu et al. [10] used the same technique to model the cities of London and Manchester with stochastic geometry. The same traffic distribution can be also observed in the public data from Google compute clusters [26, 27] and in the daily population in Manhattan [28]. Unfortunately, huge compute server clusters normally cover extensive areas and the workload variations do not represent the small scale network variations that can be found in the RAN. This can be observed in Manhattan's population [28], where different neighborhoods have different distributions.

3. Scenario Under Evaluation

The target scenario is represented by a metropolitan area consisting of differentiated zones, namely: residential, commercial and manufacturing. The deployed RAN infrastructure consists of mBSs and pBSs, which provide connectivity to both the static and mobile devices. The mBSs, which belong to the already deployed infrastructure, are stationary and have higher transmission power than the pBSs, thus covering a wider range. In contrast, the pBSs are mobile, with shorter coverage ranges, that can eventually be relocated at any time of the day based on the traffic demands of a particular region. For this, one may consider, for example, a drone carrying a pBS [29]. In this described scenario we assume that some legacy infrastructure, i.e. the mBSs, is unable to support the peaks in traffic load caused by user mobility dynamics and we require the eventual deployment of pBSs over certain regions to guarantee a good level of service. This scenario is further supported by [2] highlighting that, by 2022, an important percentage of network connections will originate from devices on the move. The node's mobility affect directly to the amount of traffic that the network will have to support along the day. The coverage of the base stations is limited and the users will connect to the closest base station. If many users move to the same location (for example, because they commute to work), the base stations can be overloaded and they may not support the requested traffic.

In the described scenario we consider a hardware disaggregation in both the mBS and pBS into a set of distributed Remote Radio Heads (RRHs)¹, which bear the lower layer processing, and a centralized Baseband Unit (BBU) pool addressing upper layer and RRM tasks. This RAN architecture is commonly referred to as Centralized-RAN (C-RAN). In this paper, we consider two options for the placement of the BBU pool, namely, in the edge network (termed herein as Edge Computing) or in the core network (termed as Cloud Computing). Fig. 1 shows a graphical representation of the scenario under evaluation. The fronthaul and backhaul of the scenario are completely wireless as the one proposed by the Third Generation Partnership Project (3GPP) in [30]. The connection between the RAN and the core network takes place through microwave aggregation switches. The aggregation switches multiplex and forward the traffic to the server (containing the

¹For convenience, and with a certain notation abuse, we will still refer to mBS and pBS instead of mRRH and pRRH in the remainder of the paper.

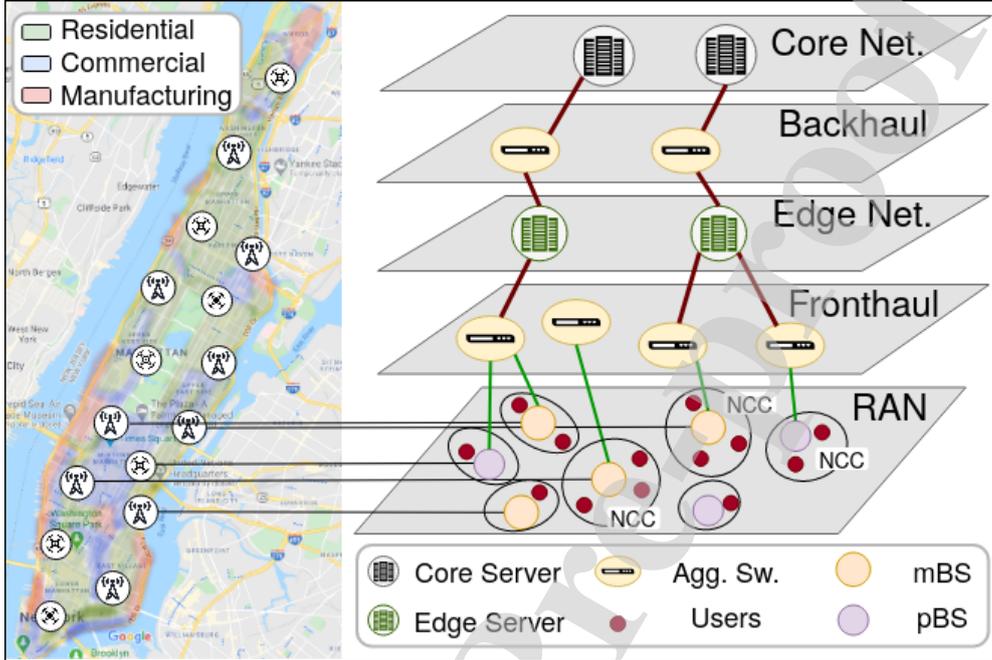


Figure 1: Graphical representation of the scenario under evaluation.

BBU pool), which is either located in a big data center in the core network or a small data center at the edge of the network.

The mobility in the proposed scenario is modeled as a random Poisson Point Process (PPP) that calculates the distribution of users in the whole area. Towards that end, we take snapshots of the area every hour and derive the likelihood for a user to be in a certain position. Then, we assume that the users can move either outside the delimited UZ zone or within their zone, based on a certain probability and on an estimation of traffic data taken from the population of the city of New York [28].

3.1. Stochastic Geometry for Functional City Zoning

To model the aforementioned scenario, we resort to the field of graph theory and leverage on the proposed framework by Bassoli et al. [9]. Specifically, the authors propose the usage of multilayer random hypergraphs, which is a graph's generalization which allows edge connections of groups of nodes (instead of connecting a single pair of nodes). Then, this abstraction is used to model a wireless cellular network as the one previously described in Fig. 1.

Let $H = (X, E)$ be the planar hypergraph representing the physical network, where X is the set of nodes, and E is the set of hyperedges. Then, let $\mathcal{H} = (X, E, X_{\mathcal{H}}, E_{\mathcal{H}}, L)$ be an extension of H , which is the random multilayer hypergraph where:

- X is the set of random nodes
- E is the set of random edges
- L is the set of layers
- $X_{\mathcal{H}}$ is the set of node-layer elements
- $E_{\mathcal{H}}$ is the set of edge-layer elements

The set X is divided in multiple subsets $X = \{X_{mBS}, X_{pBS}, \dots\}$. Each subset represents the nodes with the same function. In particular,

- X_{mBS} represents the mBS
- X_{pBS} represents the pBS
- X_{sw} represents the aggregation switches
- X_{edge} represents the edge nodes
- X_{core} represents the core nodes
- X_{UE} represents the users

The sets X_{mBS} , X_{sw} , X_{edge} , and X_{core} are modeled as PPP with intensities λ_{mBS} , λ_{sw} , λ_{edge} , and λ_{core} respectively. The set of users X_{UE} is also modeled as a PPP with a variable intensity λ_{UE} that depends on the traffic density T . Moreover, to model user's mobility, the set X_{pBS} is divided into two subsets: the mobile users and the static users. We impose a hard restriction on the mobile users: They cannot connect to the pBS because the pBS coverage is small and the users would constantly trigger handovers between BSs. The set X_{pBS} , the set of pBSs, has a variable intensity λ_{pBS} . Two different RRM heuristics are proposed to determine the density of X_{pBS} : a random model (see subsection 8.1) and a smart model (see subsection 8.2). The BSs (both the mBSs and the pBSs) are connected to the nearest aggregation switch in

the fronthaul to minimize the propagation delay. The aggregation switches are homogeneous, and thus, they form a Voronoi tessellation. Consequently, each BS of the j -th tier has an average load of $\frac{N_j}{\lambda_j}$, where N_j is the average fraction of users served by the j -th tier. To minimize the power consumption, we assume that the BSs are connected to the closest aggregation switch, which means that each aggregation switch serves the BSs inside its Voronoid cell. Di Renzo et al. [31] formulated the Probability Mass Function (PMF) of the number of BSs connected to an aggregate switch as:

$$P[N_{BS} = n] = \frac{3.5^{3.5} \Gamma(n + 3.5) (\lambda_{BS}/\lambda_{sw})^n}{\Gamma(3.5) n! (\lambda_{BS}/\lambda_{sw} + 3.5)^{n+3.5}}, \quad (1)$$

where λ_{BS} is the sum of the intensities of all BSs:

$$\lambda_{BS} = \lambda_{pBS} + \lambda_{mBS}, \quad (2)$$

and $\Gamma(x)$ is the gamma function. We obtain the average fraction of users served by the j -th tier of base stations, for both the random model and the smart model. These models are described in subsections 8.1 and 8.2.

3.2. A Primer on Network-coded Cooperation

One of the main objectives of NCC is the massive and real time transmission of data to co-located users in cellular networks. The UEs require short-range communication links that enables the cooperation among them. Therefore, the implementation of MSCs by means of deploying nearby pBS facilitates the use of NCC. In brief, NCC is enabled by three entities [17]: a content server, a controller, and the nodes forming the MSCs. The NCC content server is in charge of encoding the information and transmitting it to the MSCs. The NCC controller is in charge of organizing, creating and dismantling the MSCs. In addition, in our considered scenario, the controller is in charge of positioning the pBS to their optimal location (more on this in section 8). Fig. 2 shows the target scenario and the agents of NCC. More information about NCC can be found in Chapter 3 of [32].

In NCC, the content server divides the content into data blocks, named generations, resulting in g packets. Furthermore, each UE has a coding matrix of size g to store the packets of the generation. In the encoding process, the encoder multiplies the source packets by a vector of random coefficients, thus generating a coded packet, which is a linear combination

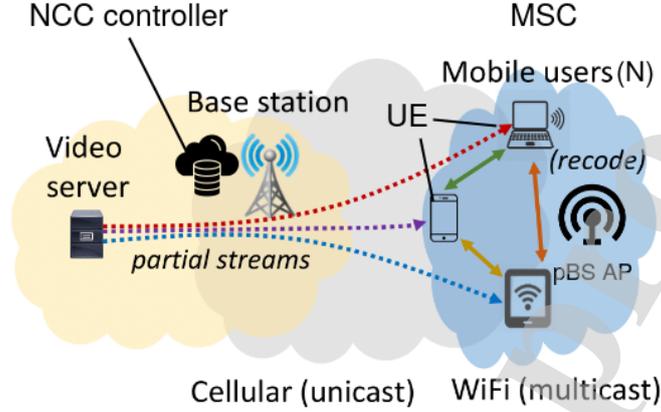


Figure 2: Concept of Network-Coded Cooperation [17].

of the packets in the generation. In the decoding process, the coefficients of the packets received are stored in the coding matrix. The rank of the coding matrix is defined as the number of linearly independent rows in the coding matrix. When the rank of the coding matrix is the same as the generation size g , the UE is able to decode the generation.

Each generation is transmitted to the UEs in two phases, namely the cellular phase and the MC phase. These two phases can happen simultaneously, if the UEs supports Multi-connectivity (MCo), or sequentially, using Single-connectivity (SCo). Noteworthy, in the case of MCo, if some packets get lost during the broadcast phase, the UEs will have to wait until the errors are corrected. The two NCC phases are briefly described hereafter:

1. **Cellular Phase:** We consider the cellular channel can be multiplexed in either time or frequency. At the beginning of the broadcast phase, each UE receives an index number, namely the UE identifier (UEID). Then, the server distributes the content to every UE inside the MSC in a round robin fashion, starting with the UE with the lowest UEID. At the end of this phase, each UE receives around $(g + c)/n$ packets (depending on the UEID), where g is the generation size, c is the extra coded packets added per generation, and n is the number of UEs in the MSC.
2. **MC Phase:** The relaying phase can run agnostic to the cellular network. In this phase, the UEs distribute the packets received from the

broadcast phase and they eventually decode the generation when the rank of the coding matrix reaches g . The pBS creates a Time-Division Multiplexing (TDM) schedule in which each UE is assigned a timeslot, which depends on the previous index assigned by the NCC controller. In each timeslot, a different UE multicasts a packet to the multicast group in the MSC. By doing so, the resources in the MSC are distributed uniformly. It is worth noting that the timeslots in this phase do not need to have the same length as in the broadcast phase since data rates from both phases may be different.

At the end of the MC phase, each UE has received at least g linearly independent packets, which enables the UE to decode the generation. If the number of linearly independent packets received by the n -th UE is $g_n < g$, the generation is said to be partially decoded.

The use of NCC results in less power consumption for both the UEs and the cellular Access Point (AP), higher downlink throughput, and cellular traffic offload. In previous works [16, 17], we have presented the impact of NCC on the aforementioned metrics for different configurations. In particular, this article leverages the analysis of the power consumption reduction as the MC size, which refers to the user density in the MC, changes.

4. Urban Zone Traffic Density Distribution

The UZ model for the metropolitan area divides the city into three different zones, namely: residential, commercial and manufacturing zones. Unfortunately, traffic models characterizing these zones are missing in the literature. Therefore, in this section we provide some estimations based on population density and mobile device penetration studies.

The city of New York has recently published several reports [33, 34, 28, 35] that allows the calculation of the population density for each of the zones during weekdays. In particular, [35] helps to identify and divide the urban zones into groups. Then, [33], [34], and [28] helps to estimate the population in the different zones and to extend it to the New York scenario. Consequently, the mean value provides reliable data of the population distribution in the three zones separately. It was observed that the population distribution during each weekday is similar to each other and, likewise, the population distribution during each of the weekend days. Hence, the population density can be

averaged over weekdays and weekend days and summarized in the population density over the weekdays and weekends, respectively.

Let $\mathbb{A} = \{\text{residential, commercial, manufacturing}\}$ be the set of zones in the urban area. Let $\delta_a(t)$ be a Random Variable (RV) of the population density per square kilometer (km^2) of the zone $a \in \mathbb{A}$ at a given time of the day t . Then, let $T_a(t)$ be a RV that defines the traffic intensity per km^2/hour in the zone a . $T_a(t)$ depends on the population density of the zone a , $\delta_a(t)$, the number of devices per capita c , and the average traffic a device generates per hour r_{hour} . Then, $T_a(t)$ is calculated as:

$$T_a(t) = \delta_a(t) \cdot c \cdot r_{hour} \quad \forall a \in \mathbb{A}. \quad (3)$$

The number of devices per capita and the monthly traffic generated by the different connected devices are extracted from Cisco's reports from 2017 [2] and 2018 [36]. In our scenario, these values are 13.4 and 1.4GB, respectively. This allows the calculation of r_{hour} as the average monthly traffic divided by the number of hours per month. Let $\mathbb{D} = \{\text{tablets, PCs, TVs, Smartphones, Machine-to-machine type communications (M2M), others}\}$ be the set of connected devices included in Cisco's report [2, 36]. The average traffic of a device can be calculated as:

$$r_{hour} = 1/|\mathbb{D}| \cdot \frac{\sum_{d \in \mathbb{D}} N_d \cdot T_d}{h_{month}}, \quad \forall d \in \mathbb{D} \quad (4)$$

where $|\cdot|$ represents the number of elements in \mathbb{D} , N_d is the percentage of connected devices of type d , T_d is the average monthly traffic generated from a device of type d , and h_{month} is the number of hours per month. The distribution of the number of connected devices is very likely to depend on the city zones (for example, the manufacturing zone is expected to have more M2M communications than the residential zone). This would imply that N_d depends on the zone a , i.e. $N_d(a)$. However, no references that could defend this statement were found in the literature and thus, we will assume equal percentage of devices of type d irrespective of the considered zone.

Finally, by combining (4) into (3), we obtain the traffic intensity distribution over a weekday and weekend day, which is plotted in Fig. 3. In it, we observe that the traffic in the residential area increases during the morning (before people go to work) and in the evening (after people go to work). On the other hand, the traffic in commercial and manufacturing areas increases during working hours. This behavior is expected since commercial and man-

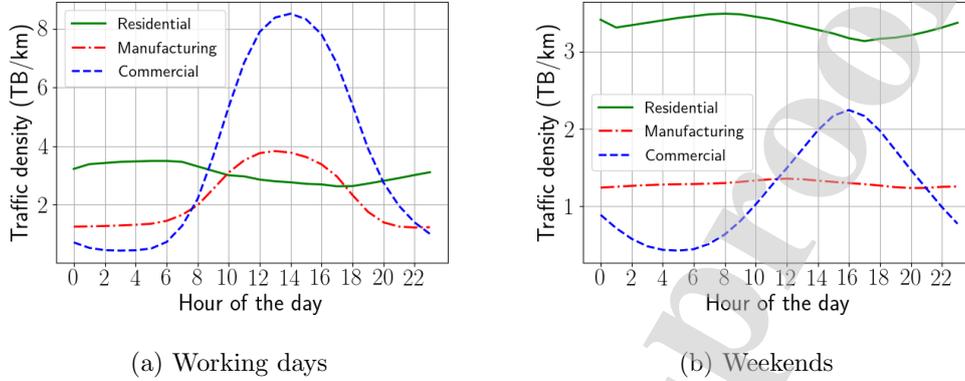


Figure 3: Hourly distribution of traffic density over the day.

ufacturing areas are expected to have more population (and therefore, more traffic) during working times. The traffic peak in commercial areas is higher than in manufacturing areas. This is also expected since commercial areas normally consist of huge skyscrapers with lots of offices for many people while manufacturing areas contain also wide industrial warehouses and factories.

When comparing zone by zone, we note that the traffic intensity variations during weekends are negligible in comparison to the traffic density variations during weekdays. Consequently, we focus on the worst case scenario, which takes place during the weekdays.

5. Power Consumption Model

This section encompasses the power consumption models used for the UZ scenario proposed in section 3. Most of the models are adopted from the EARTH european framework FP7-ICT [37]. In this work, we particularly focus on the downlink since the main focus of NCC is downlink data dissemination. Let P_{tot} be the total power consumption in the network, obtained as the sum of the individual power consumed in every element of the network, hence:

$$P_{tot} = P_{RAN} + P_{sw} + P_{edge} + P_{core}, \quad (5)$$

where P_{RAN} , P_{sw} , P_{edge} , and P_{core} are the power consumed in the RAN, in

the aggregation switches, in the edge nodes, and in the core nodes. It should be noted that the power consumed by the UEs is negligible in comparison to the power consumed by the network [38] and it is therefore not considered in this work. Moreover, the power consumed scales with the number of devices so the total power consumed depends on the extension of the territory that we want to cover. In order to have comparable metrics we use density values. For example, instead of the total power consumed from mobile base stations we use the power consumed density per square kilometer.

5.1. Power consumed in the RAN (P_{RAN})

The power consumed in the RAN is calculated as the sum of the power consumed in the corresponding base stations (pBS and mBS), taking into account that they only include the RRH, as we discussed in section 3. P_{RAN} is calculated as:

$$P_{RAN} = \sum_{j=0}^{N_{mBS}^{tot}} P_{j,mBS} + \sum_{i=0}^{N_{pBS}^{tot}} P_{i,pBS}. \quad (6)$$

where N_{mBS}^{tot} and N_{pBS}^{tot} are the total number of mBS and pBS in a square kilometer and $P_{j,mBS}$ and $P_{i,pBS}$ is the power consumed by the j -th mBS and i -th pBS, respectively. However, due to the PPP distribution of users on average it can be assumed that $P_{j,mBS} = P_{mBS}$ for every $j = \{0, 1, \dots, N_{mBS}\}$ and $P_{i,pBS} = P_{pBS}$ for every $i = \{0, 1, \dots, N_{pBS}\}$. Hence, equation (6) can be simplified to

$$P_{RAN} = N_{mBS} \cdot P_{mBS} + N_{pBS} \cdot P_{pBS}. \quad (7)$$

where N_{mBS} and N_{pBS} are the density of mBS and pBS in a square kilometer. Furthermore, the consumed power model for a base station $x \in \{m, p\}$ can be linearly modeled depending on the load [39]:

$$P_{xBS} = (1 - \rho_{xBS})P_{xBS}^{idle} + \rho_{xBS}P_{xBS}^{max} \quad (8)$$

where P_{xBS}^{idle} and P_{xBS}^{max} are the power consumption of the base station x in idle and maximum load, respectively. Furthermore, pBSs allow the use of NCC due to their short range, assuring small distance between the users connected to a specific pBS. Consequently, the power consumed in the pBSs (P_{pBS}) is at the same time multiplied by the energy reduction factor $\overline{E_{NCC}(\lambda_{pBS})}$ [17, 40].

5.2. Power consumed in the aggregation switches (P_{sw})

The power consumption of the aggregation switches can be decomposed into the consumed power during internal switching during mux/demux processes and the consumed power of the transmission/reception over their input/output links. The fronthaul and backhaul considered in this study are assumed to be wireless, that is, the aggregation switches operate over wireless interfaces. The power consumption in the aggregation switches scale with the density of base stations (N_{BS}), the load of each base station n (ρ_n), the number of antennas (N_{ant}^n), the wireless link transmission power (P_{link}^n), and the consumed power of the switches due to the mux/demux processes (P_{switch}^n) [41]:

$$P_{sw} = N_{BS} \cdot P_{switch}^n(\rho_n) + N_{ant}^n P_{link}^n(\rho_n), \quad (9)$$

where ρ_n corresponds to the aggregated load at the switch normalised by the maximum load.

5.3. Power consumed in the edge and core nodes (P_{edge} , P_{core})

The edge and core nodes can each be represented as a 3-tier network. The edge nodes comprises aggregation switches and servers where aggregation switches route the traffic to the servers, being the leaf nodes. The power consumption is the sum of both component nodes [42] and is expressed as follows

$$P_{edge} = \sum_{N_{esw}} P_{esw} + \sum_{N_{es}} P_{es}, \quad (10)$$

where P_{esw} and P_{es} is the power consumed in the edge switches and servers respectively. Similarly, the core nodes can be represented by a 3-tier network consisting of core switches and core servers. The power consumption of the core nodes is given by

$$P_{core} = \sum_{N_{csw}} P_{csw} + \sum_{N_{cs}} P_{cs}, \quad (11)$$

where P_{csw} and P_{cs} are the power consumed in core switches and core servers respectively. The power consumption follows a linear dependency to the load in the switches and the servers of both the edge and core networks. This dependency, in the edge network, is given by

$$P_{esw} = (1 - \rho_{esw})P_{esw}^{idle} + \rho_{esw}P_{esw}^{max}, \quad (12)$$

$$P_{es} = (1 - \rho_{es})P_{es}^{idle} + \rho_{es}P_{es}^{max}, \quad (13)$$

where ρ_{es} is the load of the edge network, P^{idle} represents the consumed power in idle and P^{max} represents the consumed power with maximum load. Then, in the core network, this dependency is given by

$$P_{csw} = (1 - \rho_{csw})P_{csw}^{idle} + \rho_{csw}P_{csw}^{max}, \quad (14)$$

$$P_{cs} = (1 - \rho_{cs})P_{cs}^{idle} + \rho_{cs}P_{cs}^{max}, \quad (15)$$

where ρ_{cs} is the load of the core network, P^{idle} represents the consumed power in idle and P^{max} represents the consumed power with maximum load.

6. Fronthaul and Core Network Restrictions

6.1. Computation effort per UE and per BBU

The power consumption at the server can be estimated based on its load, as the sum of the requested computation effort by all users [42], or the Giga-operations per Second (GOPS) per user [43]. In this paper the computation effort is based on GOPS, E_u per user u , and given by

$$E_u = (3A_u + A_u^2 + \frac{1}{3}M_u C_u L_u) \cdot \frac{R_u}{10}, \quad (16)$$

where A_u is the number of used antennas, M_u the modulation bits, C_u the code rate, L_u the number of spatial MIMO-layers and R_u the number of assigned Physical Resource Blocks (PRBs). The number of assigned PRBs is variable and depends on the requests of each user and the resources available in the network.

Now, let \mathbb{B} be the set of BBUs and $U_b^{BBU} \subseteq X_{UE}$ be the subset of users that request resources from the b -th BBU. For simplicity, it is assumed that every user generates the same computation effort to the BBU. Consequently, the overload of the j -th BBU, hereafter denoted as O_j , can be defined as:

$$O_b^{BBU} = \max(0, U_b^{BBU} \cdot E_u - C_b) \quad (17)$$

where C_b is the capacity (in GOPS) of the b -th BBU.

6.2. Fronthaul Capacity

The capacity of the fronthaul link in the base stations supposes an important bottleneck since many users coexist in a small area. The data rate (in bps) generated by each user depends on the coding ratio, the modulation bits, the number of antennas, and the number of PRBs assigned to that user, which in 5G New Radio (NR) is given by

$$DR_u = N_{sym} \cdot N_{sub} \cdot C_u \cdot M_u \cdot R_u \quad (18)$$

where N_{sym} , N_{sub} , C_u , M_u , and R_u are the number of symbols, the number of subcarriers, the coding rate, the modulation bits, and the number of PRBs, respectively. The aggregated fronthaul bandwidth required by the users that are served by the n -th base station is defined as

$$R_{FH,n} = \sum_{u=0}^{X_{UE}} b_{u,n} \cdot DR_u, \quad (19)$$

where $b_{u,n}$ is a binary variable equal to 1 when user u is served by the n -th base station and 0 if not. Assuming a linear correlation between the consumed power and the load of the fronthaul link [44], the power consumed by the fronthaul link of the n -th base station is given by

$$P_{FH,n} = \frac{R_{FH,n}}{C_n} \cdot P_{FH,n}^{max} \quad (20)$$

where $P_{FH,n}$, $R_{FH,n}$, C_n , and $P_{FH,n}^{max}$ are the power consumption, the aggregated data rate, the maximum capacity, and the maximum power consumed in the fronthaul link towards the n -th base station.

Now, let U_n^{BS} be the set of users that are served by the n -th base station. Then, assuming that every user generates the same traffic, the overload of the n -th base station, hereafter denoted as O_n^{BS} , can be defined as

$$O_n^{BS} = \max(0, R_{FH,n} - C_n) \quad (21)$$

where $R_{FH,n}$ is the aggregated fronthaul data rate of the n -th base station.

7. Radio network parameters and models

The path loss models are used to calculate the average coverage per base station. The path loss models for microcells (PL_m) and picocells (PL_p) are defined in the 3GPP TR 138 901 R14 [45] as

$$PL_p = 32.4 + 17.3 \cdot \log_{10}(d) + 20 \cdot \log_{10}(f) \quad (22)$$

$$PL_m = 32.4 + 21 \cdot \log_{10}(d) + 20 \cdot \log_{10}(f) \quad (23)$$

where d is the distance from the UE to the antenna and f is the carrier frequency. In this equations, we are assuming line of sight for the pBS since they are going to be mobile and we can move them accordingly, avoiding blind spots. Moreover, regarding the mBS, it should be noted that TR 138 901 R14 [45] also defines a threshold under which the path loss model of microcells in (23) is valid

$$d_{break} = 4 \cdot h_m \cdot h_u \cdot f / c \quad (24)$$

where h_m is the height of the microBS, h_u is the height of the user, and c is the constant of the speed of light in vacuum space. For example, for $f = 5GHz$, $h_m = 32m$, and $h_u = 1.5m$, $d_{break} = 3.2km$. In UDNs it can be assumed that this model is valid since the high density of base stations has a direct implication that the range of these base stations is going to be small. TR 138 901 R14 [45] also defines a threshold stating that the mBS must have a range over 10m range. We assume that the mBS will have at least 200m range and therefore, this threshold does not affect us.

The path loss models are used to calculate the SINR, which is required to obtain the range of pBSs and mBSs. The SINR ($\gamma_p(d_j)$) a user perceives from the j -th pBS, which is located at a distance d_j , is obtained by the following equation [21]:

$$\gamma_p(d_j) = \frac{P_{out}^{pBS_j} h_{x_j} d_j^{-\alpha}}{\sum_{i=1, i \neq j}^{N_{pBS}^{tot}} P_{out}^{pBS_i} h_x d_i^{-\alpha} + \sum_{i=1, i \neq j}^{N_{mBS}^{tot}} P_{out}^{mBS_i} h_x d_i^{-\alpha} + \sigma^2} \quad (25)$$

where $P_{out}^{pBS_j}$ is the transmission power of the the j -th pBS, h_{x_j} is the Rayleigh fading between the user and the j -th pBS, α is the path loss element, and σ^2 is the constant additive noise power. Finally, the user is in coverage if:

$$\gamma_p(d_j) > \theta_j, \quad (26)$$

where θ_j is the SINR threshold for the j -th pBS. In the case that all pBSs have the same capabilities then $\theta_j = \theta \forall j \in N_{pBS}^{tot}$ is assumed. Equation (25)

states that the SINR from a user to the j -th pBS does not only depend on the distance to the j -th pBS but also on the distance to the remaining base stations in the network. Moreover, equation (25) also shows that closer base stations to the user have more impact on the SINR than the ones further away. The scenario presented has a variable number of base stations, where some of them are mobile. Unfortunately, there are no methods to estimate the SINR in an area with so many different unknowns. However, considering that the effect of further base stations is reduced to beamforming, we can assume that SINR is upper bounded by the closest interfering base station [46]. Consequently, the next objective is to derive an expression for the distance of a user to the closest base station.

Towards that end, we calculate the minimum distance between x_u (the RV with the position of the user in \mathbb{R}_2) and X_{mBS} (a set of RV with the positions of N_{mBS} mBS in \mathbb{R}_2) in a map $n - by - n$ where n is the number of tens of meters. Then, let $d\{\cdot\}$ be the operator that calculates the distance between two points in \mathbb{R}_2 . The minimum distance (d_{mBS}^{min}) between x_u and X_{mBS} is the minimum of the distances between x_u and x_{mBS} , for every $x_{mBS} \in X_{mBS}$.

$$d_{min} \triangleq \min d\{x_u, X_{mBS}\}$$

$$d_{min} = \min [d\{x_u, x_{mBS}\}], \quad \forall x_{mBS} \in X_{mBS}. \quad (27)$$

The utilization of NCC implies longer minimum distances [40]. In order to calculate the minimum distance, we developed a simulator that computes d_{min} for every value of x_u , X_{mBS} , and n . We evaluated the results of the simulator for lower values of x_u , X_{mBS} , and n to extrapolate the results to a fitting equation that only depends on d_j . The MSE due to the extrapolation process is $3.03 \cdot 10^{-6}$.

Assuming that the closest interfering base station is an mBS, the minimum distance d_{mBS}^{min} can be included into equation (25) to obtain the maximum distance d_j between the j -th pBS and the user to provide a SINR $\gamma_p(d_j) > \theta_j$. In particular, we calculate $\gamma_p(d_j)$ for the case where only the closest mBS is interfering. In a similar manner, the minimum intercell distance between pBS (d_{pBS}^{min}) can be obtained from equation (25), taking into account the distance d_j obtained from evaluating $\gamma_p(d_j)$ for d_{mBS}^{min} . As a result, we observe a linear relation between the distance of the serving pBS and the interfering mBS/pBS that is defined by

$$d_{mBS}^{min}(d_j) = 4.57d \quad (28)$$

$$d_{pBS}^{min}(d_j) = 1.27d. \quad (29)$$

It should be noted that the calculation of minimum distance takes into consideration the application of NCC to pBSs, which, based on the user density in the pBS, impacts on the power consumed [40].

8. Optimal mBS Density and pBS Density per City zone

The RAN infrastructure consists of a fixed part (the mBS) and a mobile part (the pBS). The fixed nature of mBSs implies that the mBS density calculation must be done separately for each zone. The mBS density is defined by the mBSs available by the existing RAN infrastructure which depends on the city zone. The existing mBSs set a lower bound on the mBS density. This number depends on many factors such as the maximum network traffic or the population. Instead of using the existing infrastructure, we optimize the mBS density so that the power consumed in the RAN is minimized. It should be noted that the BS density is not defined by the deployed BS that can be available but by the actual active BSs. Hence, the problem can be defined as follows:

$$\begin{aligned} & \min P_{tot}^{RAN}(T, t) \\ \min & N_{mBS} \cdot P_{mBS}(T, t) + N_{pBS}(T, t) \cdot P_{pBS}(T, t) \end{aligned} \quad (30)$$

where N_{pBS} and N_{mBS} are the pBS and mBS densities respectively, P_{pBS} and P_{mBS} is the average power consumed for pBSs and mBSs, and T is the traffic density from Fig. 3. Then, let \mathbb{U} , \mathbb{M} , and \mathbb{P} be the set of users, mBS, and pBS in the network, respectively. It should be noted that T also depends on t , the time. The minimization of $P_{tot}(T, t)$ has the following constraints:

- a) $b_{u,x} \in \{0, 1\} \quad \forall u \in \mathbb{U}, x \in [\mathbb{M} \cup \mathbb{P}]$
- b) $\sum_{u \in \mathbb{U}} \sum_{p \in \mathbb{P}} \sum_{m \in \mathbb{M}} (b_{u,p} + b_{u,m}) = 1$
- c) $\sum_{u \in \mathbb{U}} \sum_{p \in \mathbb{P}} \sum_{m \in \mathbb{M}} (b_{u,p} \cdot b_{u,m}) = 0$

$$d) P_{pBS}(T, t) = P_{pBS}^{mob} + (1 - \rho_{pBS})P_{pBS}^{idle} + \rho_{pBS}P_{pBS}^{max}$$

$$e) P_{mBS}(T, t) = (1 - \rho_{mBS})P_{mBS}^{idle} + \rho_{mBS}P_{mBS}^{max}$$

$$f) \rho_{pBS}(t) = \frac{r_{NCC}(\lambda_{pBS}) \cdot T_{pBS}(t)}{N_{pBS} \cdot C_{pBS}}$$

$$g) \rho_{mBS}(t) = \frac{T_{mBS}(t)}{N_{mBS} \cdot C_{mBS}}$$

$$h) P_{pBS}^{mob} = (c1 \cdot (w_{UAV} + w_{battery} + w_{ant}) \cdot c2) \cdot t_{air}$$

$$i) C_{mBS} = C_{pBS} = C_{BS}$$

$$j) N^{max} = \frac{T^{max}}{C_{mBS}}$$

$$k) N_{mBS} \leq N^{max}$$

$$l) N_{mBS}^{min} = N_{mBS} \mid \int_t^{t+1} P_{mBS}(T, t) dt \leq \Delta_{mBS}^P(0)$$

$$m) N_{pBS}^{min} = 0$$

where $b_{u,x}$ is a binary variable equal to 1 when user u is served by base station x . Constraint (a) establishes a binary decision. Constraints (b) and (c) establish that a user must be connected to one and only one base station. Constraints (d) and (e) follow the energy consumption model from section 5.1. The constraints are an adaptation from constraint (8). Constraints (f) and (g) represent the average load for pBSs and mBSs. The factor $r_{NCC}(\lambda_{pBS})$ represents the increase of supported load due to NCC and it depends on the average user density in the pBSs. NCC is not applied in mBS and thus, the term $r_{NCC}(\lambda_{pBS})$ is not present in constraint g). Constraint (h) shows the mobility cost in terms of power consumption of the pBS, considering that the mobile pBSs are located on UAVs. The values are taken from a linear regression of an experimental evaluation from Tseng et al. and the coefficients $c1$ and $c2$ have the values 160 and -60 , respectively [47]. Then, w_{UAV} , $w_{battery}$, and w_{ant} are the weights of the UAV, an external battery to increase the flying time of the UAV to one hour, and the pBS antenna. Finally, t_{air} is the time it takes for the UAV to move between designated locations. Constraint (i) indicates that the mBS and pBS

capacity is equal and with a constant value C_{BS} . Constraint (j) calculates the maximum density of base stations based on the maximum traffic and the capacity of the base stations. Constraint (k) establishes an upper bound for the maximum density of mBS. Constraint (l) indicates that the minimum mBS density corresponds to the mBS density such that a new mBS would decrease the utilization of the mBS. Moreover, constraints (l) and (m) establish a lower bound for the number of mBS and pBS.

The minimization problem is evaluated for $N_{min}^{mBS} \leq N^{mBS} \leq N^{max}$ along the hours of the day. As a result, equation (30) can be redefined as a function of t and N_{mBS}

$$P_{tot}^{RAN} = f(t, N^{mBS}). \quad (31)$$

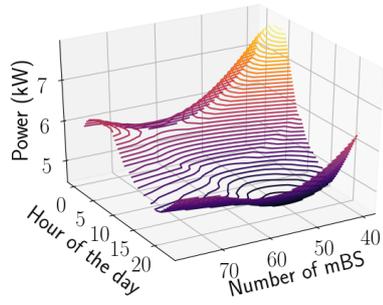
The objective is to minimize P_{tot}^{RAN} with a constant number of N_{mBS} , hereafter denoted as the optimal number of mBS (N_{mBS}^{opt}) to minimize the power consumption. This is given through the integration of P_{tot}^{RAN} over t , defined by

$$N_{mBS}^{opt} \triangleq N_{mBS} \mid \int_t P_{tot}^{RAN}(t, N^{mBS}) dt, \quad (32)$$

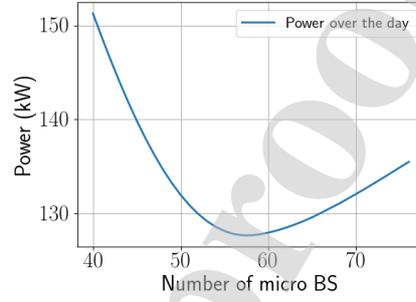
which concludes the characterization of N_{mBS}^{opt} . The traffic generated per user is 262.688 MB/h and the number of supported users per BS is 171. In the residential area, there are 59 fixed mBS/km² out of 78 BS/km². In the commercial area, there are 132 fixed mBS/km² out of 190 BS/km². In the manufacturing area, there are 61 fixed mBS/km² out of 86 BS/km². Fig. 4 (right side) shows the results of equation 32 for the scenario proposed.

Next, the optimization problem to find the minimum pBS density to minimize the power consumption in the RAN and provide the necessary network resources to support the network traffic peaks is investigated. On the contrary to the optimization of N_{mBS} , the mobility nature of pBSs enforces to find the optimal pBS density as a joint optimization problem for the three zones together. Consequently, the objective function can be described as

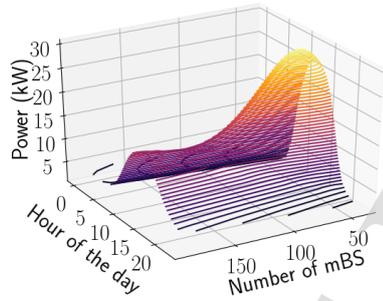
$$\min_{a \in \mathbb{A}} P_{RAN} \text{ in } \mathbb{A} \sum_{a \in \mathbb{A}} \left[N_{mBS}^a \cdot P_{mBS}^a(T, t) + N_{pBS}^a(T, t) \cdot P_{pBS}^a(T, t) \right] \quad (33)$$



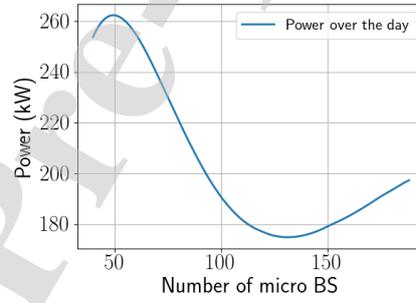
(a) Power map (Residential)



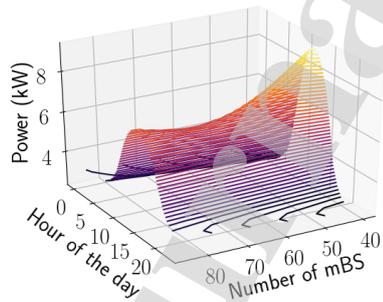
(b) Cumulative power (Residential)



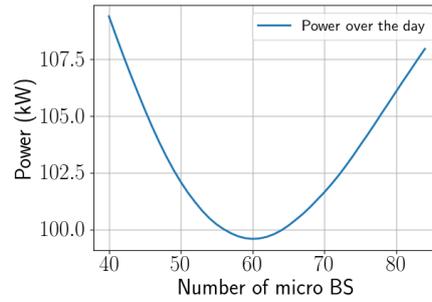
(c) Power map (Commercial)



(d) Cumulative power (Commercial)



(e) Power map (Manufacturing)



(f) Cumulative power (Manufacturing)

Figure 4: 3D power maps and aggregated power for residential, commercial, and manufacturing zones for different N_{mBS} .

The constraints in equation (33) are also applied in equation (30). In addition, the pBS density needs to be high enough to cover the traffic surplus that was not covered by the mBS. Consequently, equation (33) has the constraints (a)) - (m)) plus the following:

$$\text{n) } O_b^{BBU} = 0 \quad \forall b \in \mathbb{B}$$

$$\text{o) } O_m^{BS} = 0 \quad \forall m \in \mathbb{M}$$

$$\text{p) } O_p^{BS} = 0 \quad \forall p \in \mathbb{P}$$

where constraint (n) establishes that any BBU can be overloaded and constraints (o) and (p) establish the maximum bandwidth requested to the mBS and pBS, respectively. Constraint (33) is solved by using two different RRM heuristics explained in the following subsections.

8.1. RRM model: Random pBS Placement

In the random model, the pBS are placed randomly as a PPP. Since the transmit power of the base stations is heterogeneous, the coverage is modeled as a weighted Voronoi tessellation. Dhillon et al. [21] proposed a method to obtain the fraction of users N_j connected to base stations of type j , if all base stations and users are modeled as PPP, and some users can only connect to a subset of the set of base stations. N_j has the following distribution

$$N_j = \begin{cases} \frac{\lambda_j P_{trx,j}^{2/\alpha} \theta_j^{2/\alpha}}{\sum_{i=1}^2 \lambda_i P_{trx,i}^{2/\alpha} \theta_i^{2/\alpha}} & j \in \mathcal{B}, \\ 0 & \text{otherwise.} \end{cases} \quad (34)$$

where $P_{trx,j}$ is the power to transmit, θ_j is the SINR threshold of the base station of the j -th type, α is the path loss element, and \mathcal{B} is the subset of the set of j -th type of base stations to which the UE can be connected to.

The random model does not allow for any improvements in the position of the pBS or on the optimization of the number of users a pBS can handle. Consequently, there are no restrictions on the minimum or the maximum number of connected users to pBS apart from physical restrictions, such as sufficient SINR to receive the signal or sufficient available capacity. Algorithm 1 contains a pseudocode of the random model.

```

while Total BBU overload  $O_b^{BBU} > 0$  do
  | Set random pBS in map;
  | Get new BBU overload;
end

```

Algorithm 1: Pseudocode for the random model

8.2. RRM model: Smart pBS Placement

In the smart model, the pBS are placed in the best position to cover as many users as possible. In other words, there is an entity that controls the position of the pBS and tells them where to move at any time. This entity, named placement controller (PC), periodically monitors the traffic in the area and moves pBS to the optimal position to cover the maximum number of users. This increases the user density inside the pBS, which decreases the number of pBS and, in the end, improves the efficiency of the system.

The controller monitors the network and places the pBSs one by one. Let Λ_k be a RV that defines the user expectation of the k -th pBS placed. The users are distributed randomly with intensity λ_{UE} around the map with size $N \times N$ and area $A_{tot}(N)$. Due to the stationary property of the PPP, there is a linear correlation between the intensity λ_{UE} and the area under evaluation. Finally, let A_{pBS} be the area covered by a pBS. The PMF of the user expectation in the first pBS is

$$\Pr[\Lambda_0 = \lambda_0] = \lambda_{UE} \cdot m \cdot \frac{A_{pBS}}{A_{map(N)}}, \quad (35)$$

where m is a RV that models the probability of a user to be mobile. Building on this, the $k + 1$ -th pBS is placed in the next best optimal place after the k -th pBS and the expectation $\Lambda_{k+1} < \Lambda_k$. The difference between Λ_k and Λ_{k+1} can be estimated as

$$\Lambda_{k+1}(\lambda) = \Lambda_k(\lambda - \lambda_k) \quad \text{for } \lambda \leq \lambda_{UE}. \quad (36)$$

Moreover, the area covered by the k -th pBS is affected by the number of mBS and pBS. The impact of the number of mBS and pBS is modeled in section 7, which is used to obtain the average radius of the pBS, and hence, the average area of the pBS A_{pBS} . The higher the number (or density) of pBSs, the smaller the A_{pBS} and the number of pBS changes in each iteration. However, for simplicity of the model, this number has been reduced

to the factor $1/d_{\text{pico2pico}}$, which is the ratio between the pBS distance and the pBS minimum intercell distance to provide an SINR $\gamma_p(d_j) > \theta_j$. Building on this reasoning, the distribution of Λ_k can be defined as

$$\Pr[\Lambda_k = \lambda_k] = (\lambda_{UE} \cdot m - \sum_{i=1}^{k-1} \lambda_i) \cdot \frac{(1/d_{\text{pico2pico}}) \cdot A_{\text{pBS}}}{A_{\text{map}(N)}}. \quad (37)$$

The probability from (37) gives the user expectation of the k -th pBS placed on the map. However, the controller monitors the traffic in the network to place the pBS in the optimal position given NCC is used. Or in other words, the controller evaluates Λ_k in every potential pBS location and takes the optimal value. Consequently, let $Y_{k,n}$ be a RV with the maximum Λ_k over n occurrences defined by

$$Y_{k,n} = \max\{\Lambda_{k,1}, \Lambda_{k,2}, \dots, \Lambda_{k,n}\}, \quad (38)$$

These are various occurrences of an independent and identically distributed PPP. The Cumulative Distribution Function (CDF) of the maximum Λ_k among different occurrences of a PPP with the same expectation is calculated as the CDF to the power of the number of occurrences [48].

$$\begin{aligned} \Pr[Y_{k,n} \leq y] &= \Pr[P(\Lambda_{k,1} \leq y) \cap \dots \cap (\Lambda_{k,n} \leq y)] \\ \Pr[Y_{k,n} \leq y] &= \Pr[\Lambda_k \leq y]^n. \end{aligned} \quad (39)$$

The value of n is the number of possible locations for the new pBS taking into account the use of NCC and it depends on the map size N and the number of pBS already on the map. The value of n is estimated as

$$n(N, k) = A_{\text{tot}}(N - 2 \cdot r_{\text{pBS}}) - \sum_{i=1}^{k-1} A_{\text{pBS}}(i). \quad (40)$$

The algorithm described in 2 contains the pseudocode for this model

9. Evaluation

In this section, the evaluation of the analytic models is presented. First, we introduce the parameters and metrics selected for the evaluation. Next, we present the results of the analytic models. Then, we validate them by

```

while Total BBU overload  $O_b^{BBU} > 0$  do
  Get traffic load in all base stations;
  Estimate users from traffic load;
  MAX_pBS_LOAD  $\leftarrow 0$ ;
  foreach pBs position in potential pBS positions in map do
    Get estimated pBS load;
    if Estimated pBS load  $> MAX\_pBS\_LOAD$  then
      MAX_pBS_LOAD  $\leftarrow$  Estimated pBS load;
      Optimal position  $\leftarrow$  Current position;
    end
  end
  Calculate new pBs range;
  Calculate new average pBS density;
  Calculate new average pBS user density;
  Get total BBU overload;
end

```

Algorithm 2: Pseudocode for the smart model

means of the MSE. We compare the analytical results with the ones from a Monte Carlo simulator. Finally, we perform a sensitivity analysis on the parameters that we initially introduced at the beginning of this section.

9.1. Parameters and assumptions

The selection of parameters are collected from a wide range of works due to the extensive scope of the evaluation. Tables 1 to 4 collect the parameters used in the simulation. The mobility factor, m , ranges from 0 to 1 and corresponds to the probability that a user is mobile. For example, if $m = 0.2$ there is a 20% chance that a user is mobile and an 80% that the user is static.

9.2. Metrics

This subsection collects the metrics used to measure the performance of the models and their efficiency. The metrics used are the following:

- **Validity of the models:** The validity of the models is assessed by means of the MSE between the models and a simulator. Towards that end, a Monte Carlo simulator was built in Python.

Carrier Frequency ^[49]	f	5 GHz
Channel bandwidth ^[49]	BW	20 MHz
Number of antennas per UE ^[49]	A_u	4
Modulation bits per UE ^[49]	M_u	6
Code ratio ^[49]	C_u	0.553
Number of MIMO layers per UE ^[49]	L_u	4
User mobility factor ^[2]	m	0.5
UAV Power slope ^[49]	$c1$	160
UAV Power offset ^[49]	$c2$	-60
Weight UAV ^[49]	w_{UAV}	1800 gr.

Table 1: User and UAV parameters for simulation

SINR threshold mBS/pBS ^[21]	θ_{BS}	1
Path loss element mBS/pBS ^[21]	α_{BS}	2.7
Transmission power mBS ^[39]	P_{mBS}^{out}	6.3W
Transmission power pBS ^[39]	P_{pBS}^{out}	0.13W
Power in idle for mBS ^[39]	P_{mBS}^{idle}	15.44 W
Power in idle for pBS ^[39]	P_{pBS}^{idle}	2.46 W
Power in max load for mBS ^[39]	P_{mBS}^{max}	81.26 W
Power in max load for pBS ^[39]	P_{pBS}^{max}	7.26 W

Table 2: RAN parameters for simulation

- **User density in pBS (D_{pBS}):** The average user density increase in pBSs indicates a better utilization of the resources in the network. The user density in pBS (D_{pBS}) is defined as

$$D_{pBS} = \frac{\sum_{u \in \mathbb{U}} b_{u,p}}{N_{pBS}}. \quad (41)$$

A higher D_{pBS} means that the number of required pBS decreases, which

FH/BH antennas per mBS/pBS ^[41]	N_{ant}	2
FH/BH capacity for mBS/pBS ^[44]	C_{BS}	100 Mbps
FH/BH power max load mBS ^[44]	$P_{max,mBS}^{BH}$	50 W
FH/BH power max load pBS ^[44]	$P_{max,pBS}^{BH}$	50 W
Power region for low/high traffic ^[41]	$P_{low/high}$	37/92.5W
Traffic threshold low/high ^[41]	C_{thr}	0.5 Gbps
Switch maximum capacity ^[41]	C_{switch}	36 Gbps

Table 3: Fronthaul and backhaul parameters for simulation

Power of the edge router in idle ^[50]	P_{esw}^{idle}	120 W
Max power of the edge router ^[50]	P_{esw}^{max}	0.44 W
Numer of hops edge router ^[50]	h_e	3
Power of the core router in idle ^[50]	P_{csw}^{idle}	215 W
Max power of the core router ^[50]	P_{csw}^{max}	0.44 W
Numer of hops core router ^[50]	h_c	6
Power of the datacentre switch in idle ^[42]	P_{es}^{idle}	200 W
Power of the datacentre server in idle ^[42]	P_{cs}^{idle}	544 W
Max power of the datacentre switch ^[42]	P_{es}^{max}	300 W
Max power of the datacentre server ^[42]	P_{cs}^{max}	750 W
Server capacity ^[41]	C_b	324 GOPS

Table 4: Network and datacentre parameters for simulation

at the same time decreases the power consumed in the RAN and thus, the total power.

- **pBS density (N_{pBS}) and UAV density (N_{UAV}):** The mobility feature of pBSs enables network resources to be moved between city zones. Consequently, the pBS density depends on the zone. Moreover, the extension of each zone is not the same, which means that the contribution of required pBSs of one zone is different from the contribution

of another zone. For example, if 10 pBSs per km² are required in the commercial zone, they can split into two groups of 5 pBSs and move to two residential zones and, in the same way, two groups of 5 pBSs from two different residential zones can merge into one group of 10 pBSs and cover one commercial zone with higher pBS coverage and capacity requirements. Consequently, the pBS (and UAV) density must depend on the relative extension of the zones in the total map. As a result, the pBS density per zone is multiplied by a weight (w_a), which represents the fraction of the zone of type a in the scenario under evaluation. For the city of New York [33, 34, 28], these values are $w_r = 0.7$, $w_c = 0.2$, and $w_m = 0.1$ for the residential, commercial, and manufacturing zones, respectively. Building on this, the minimum pBS density is calculated as

$$N_{pBS}^{min} = \max \left[\sum_{a \in \mathbb{A}} w_a N_{pBS,a}(t) \right], \quad (42)$$

Similarly, the minimum UAV density is calculated by

$$N_{UAV}^{min} = \max \left[\sum_{a \in \mathbb{A}} w_a N_{UAV,a}(t) \right]. \quad (43)$$

- **Power consumed in the RAN (P_{RAN}) and total power consumed (P_{tot}):** Reducing the power consumption is one of the targets of 5G. Hence, the optimization problems were selected to minimize the power consumed. Furthermore, the pBS density, the mBS density, and the fraction of users connected to pBS and mBS significantly impact the power consumed in the RAN. P_{tot} represents the total power consumption in the network, that is, the sum of all components of equation 5. We compare the total power consumed between 4G edge computing, two approaches from the related work, and the two proposed models.
- **Sensitivity to variations in the scenario:** Although the traffic model is likely to be periodic over the week, seldom events may impact on the metrics. Examples of these events are sports events where the traffic increases or environmental catastrophes that destroy the network infrastructure. Consequently, the sensitivity benchmark also evaluates the pBS coverage and the bandwidth overload (O_m^{BS} and O_p^{BS}) as the

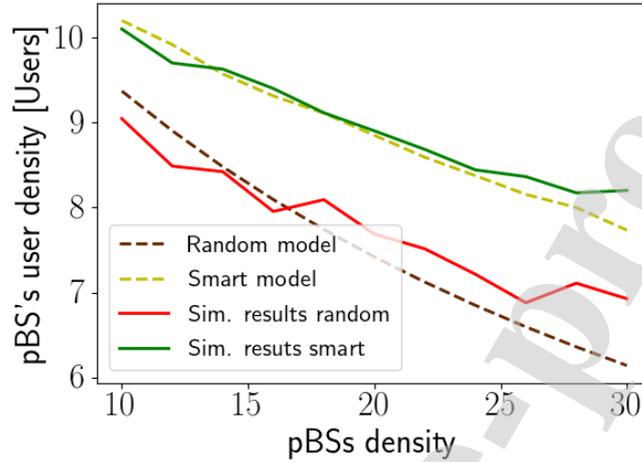


Figure 5: Comparison between the models and the simulator. The MSE between both is 0.18 for the random model and 0.03 for the smart model. $N = 500$, $\lambda = 700$, $m = 0.5$, available pico $\{10..30\}$.

ratio between resources requested and resources given, hereafter named as service efficiency.

9.3. Results

Model Validation: A Monte Carlo simulator built in Python is used to assess the validity of the mathematical models. The simulator ran a total of 50 experiments for the following configurations: $\mathbb{U} \in \{16,000..28,000\}$, $m \in \{0..1\}$, $d_{pBS} \in \{20..100\}$ meters, and $N_{pBS}^{max} \in \{10..30\}$. Fig. 5 shows the D_{pBS} for the models and the results of the simulator with $N_{pBS}^{max} \in \{10..30\}$. The MSE between the simulator and the models is 0.18 for the random model and 0.03 for the smart model.

User density per picoBS: Fig. 6 shows the distribution of users and the pBSs' user density for the random and smart models in the city zones. In the plots, the left y-axes refer to the total number of users per km^2 connected to mBSs (in blue) and pBSs (in salmon). The y-axis on the right corresponds to the user density per pBS (in brown) or, in other words, the efficiency of the model. The user density is further used in the NCC model to obtain how the efficient the model is in comparison to the use of pBS without NCC. We consider a uniform distribution of users with λ_{pBS} density in every pBS

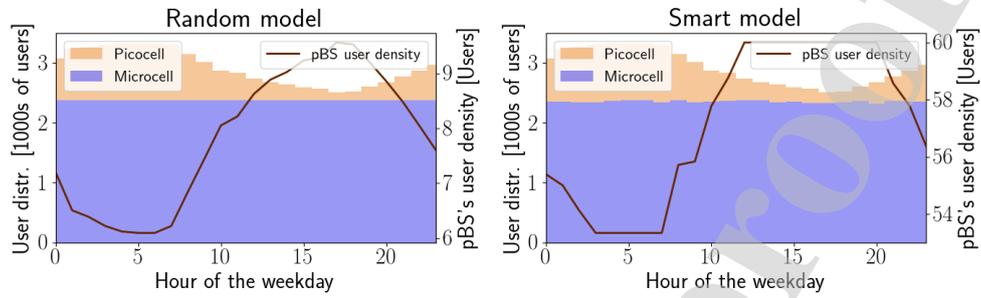
placed. It should be noted that this is an approximation of the real scenario, which would have different λ_{pBS} for different pBS. Moreover, it should be noted that the power reduction due to NCC is not linear [17].

Fig. 6a shows the distribution of users and the pBS's user density for the random and smart models in the residential zone. The small variations in traffic density (see Fig. 3) result in a constant occupation in the mBSs. In other words, the mBSs are most of the time working at 100% of their capacity. The number of pBSs changes with the traffic load over the day. Fig. 6a also shows that the pBS's user density increases the fewer the pBS density is. This is repeated in the three zones and for each of the two models. In the case of the random model, the reason for this behavior lies in the fact that the new randomly placed pBS has a chance to cover an area already covered by another pBS, and thus, the total D_{pBS} is reduced. In the case of the smart model, as observed in section 8, new pBSs when placed they always have smaller D_{pBS} than the previous ones. Fig. 6a also shows that the efficiency of the smart model increases six to eight-fold as compared to the random model.

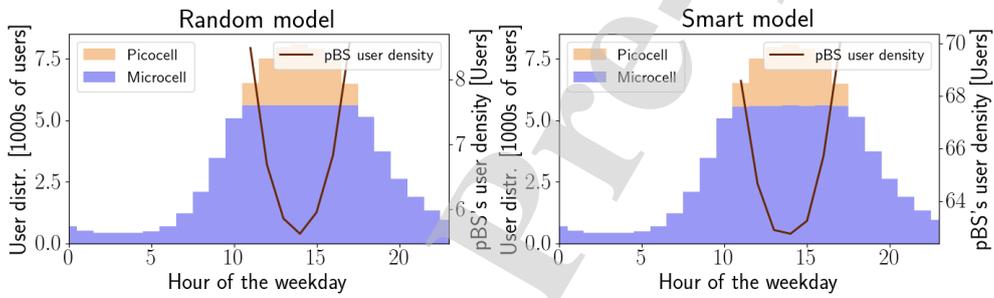
Fig. 6b shows the distribution of users and pBS's user density for the random and smart models in the commercial zone. The huge increase during working hours (see Fig. 3) impacts on the load of mBSs, which is significantly reduced during the night and it reaches 100% usage after 11 a.m. Despite the fact that the mBSs are sufficient to take care of the traffic overnight, daily traffic peaks need to be covered by pBSs. The random model does not increase the pBS efficiency in comparison to the residential zone. On the contrary, the user density for the smart model increases eight- to eleven-fold in comparison to the random model.

Fig. 6c shows the distribution of users and the pBS's user density for the random and smart models in the manufacturing zone. The evaluation in this zone is similar to the commercial zone but on a smaller scale. The traffic increase during working hours (see Fig. 3) impacts the load of mBSs. During the day, the mBSs are loaded to 100% and the pBSs take care of the traffic peaks. The random model provides a similar efficiency to the other two zones. The efficiency of the smart model increases six- to eight-fold in comparison to the random model.

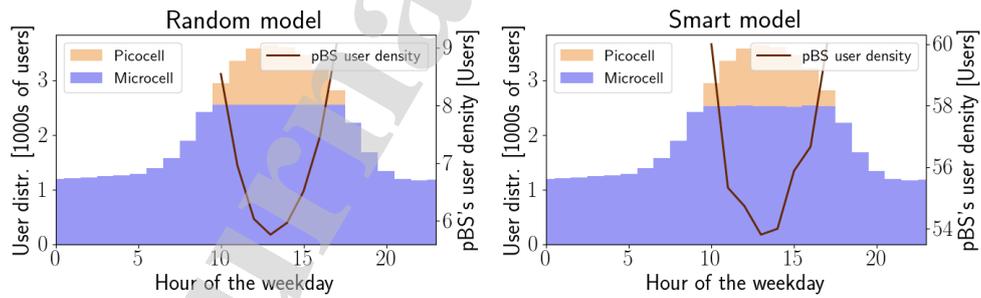
pBS density and UAV Density: Fig. 7 shows the minimum pBS density per hour of the day for the city of New York. We have compared the random and smart models with two of the approaches presented in the related works, namely the approach from Li et al. [23] and the approach from



(a) Residential zone



(b) Commercial zone



(c) Manufacturing zone

Figure 6: Evaluation of the distribution of users and the pBS's user density for the random and smart models in the city zones.

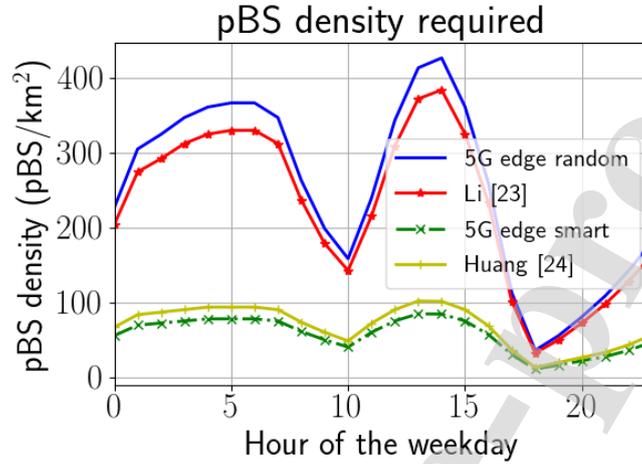


Figure 7: Minimum pBS density per hour of the day for the city of New York with the random and smart models.

Huang et al. [24]. However, during the estimation for the aforementioned approaches, we took the following estimations. First, both works take into account only the distribution of drone-BS, but they do not take into account how the previous static infrastructure is deployed. Therefore, we adapted each model so that the number of mBS was defined by the optimization strategy in section 8 and the number of pBS was defined using the models in the related works.

As expected, the pBS density required with the random model is much higher than with the smart model. There is no significant peak in pBS density during the working hours due to the fact that residential zones are much bigger than commercial and manufacturing zones. This makes sense since the population of New York does not change drastically over the day but people move inside the city (for example, to commute to work). Moreover, the smart model outperforms the model proposed by Huang et al. by around 20% and the model proposed by Li et al. by around 50 to 150%, depending on the hour of the day. The results show that the pBS density required per km^2 is 375, 102, 416, and 90 for Li's model, Huang's, and the random and smart models, respectively.

A picoBS transceiver and a battery are attached to the UAVs to enable them to fly for more than one hour. This extra devices also impact in the

energy consumed by the pBS. The pBS density is calculated assuming that one UAV can remain on the air long enough to fly from the charging position to the desired place, give coverage for at least one hour, and return to the charging position. Taking into account the values of Fig. 7 and the number of UAVs per pBS, the minimum UAV density for the city of New York is 832 UAVs/km² for the random model and 180 UAVs/km² for the smart model. If the efficiency is defined as $D_{UAV}^{random}/D_{UAV}^{smart}$, the smart model is 460% more efficient than the random model.

Power consumed: Fig. 8 shows the power consumed in the RAN, the power consumed in the fronthaul network, and the total power consumed in the proposed scenario for the residential area, the commercial area, the manufacturing area, and the complete UZ model. Moreover, the figures compare the power consumption for the current 4G LTE network with wireless fronthaul/backhaul, the two proposed models, and two of the approaches introduced in the related work.

In general, the power in the RAN with the smart model is significantly reduced in comparison to 4G LTE with edge computing. On the other hand, the massive amount of pBSs required in the random model skyrockets the power consumption in the RAN. Regarding the fronthaul network, 4G LTE consumed less power, which is attributed to the lower transmission power and smaller coverage of pBSs as compared to mBSs. As a result of this, the number of base stations required to cover the same area increases. Moreover, Fig. 8 shows a huge step in the fronthaul power consumption. The reason for that lies in the fronthaul power consumption model proposed by Sabella et al. [41], in which the authors determine three power regions of the aggregation switches. In this case, the traffic during the night falls below the threshold of one of the regions and, during the early morning, the traffic load increases enough to move to the next power region, thus increasing the power consumed in the fronthaul model. The last plot of Fig. 8 shows the total power consumed in the network for 4G LTE computing, and 5G edge computing with the random model, the smart model, and two approaches from the related work. We observe that both approaches in the related work perform better than the random model and the approach from Huang et al. is in average better. The smart model outperforms all three models achieving power savings from 6% to 25%, depending on the model and the hour of the day.

In the residential zone, the power consumed in the RAN is reduced from 13 - 16 kW to 11 - 13kW, thus achieving power savings of 17.25%. The

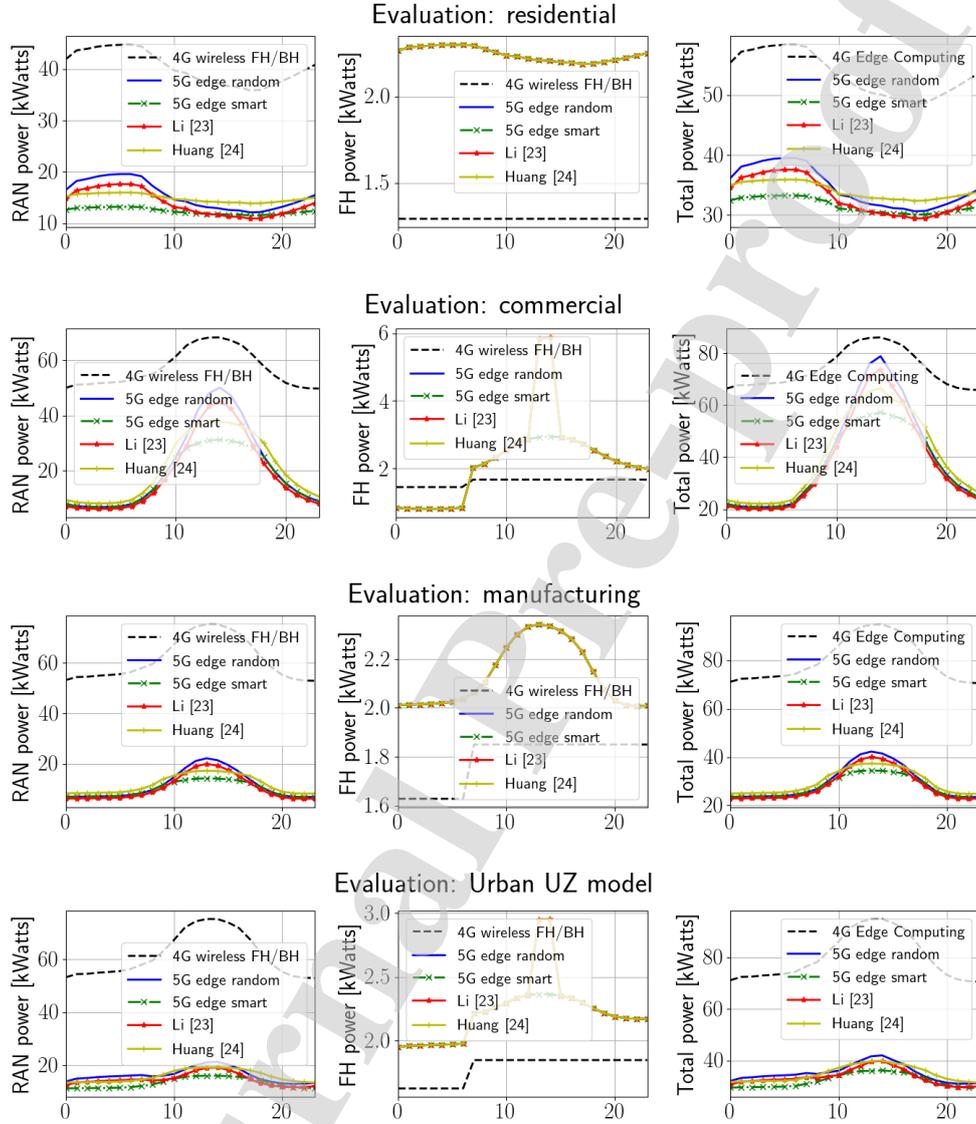


Figure 8: Evaluation of power consumed in the proposed scenario with the UZ model.

difference between 4G LTE and the new approaches is because of the fact that 4G LTE does not use mobile pBSs to cover the peaks in traffic but only mBSs. On the other hand, 4G LTE consumes less power in the fronthaul network. Regarding total power, the smart model generates around 7.7%

total power savings in comparison to the 4G edge computing approach.

In the commercial zone, the smart model achieves RAN power savings around 20% during the working hours of the day. The power evaluation also shows that the random model is much more sensitive to traffic peaks than the smart model. Regarding total power consumed, the results show power savings when using the smart model of around 14% in comparison to the edge computing approach.

In the manufacturing zone, the model shows power savings around 23% in the RAN for the smart model in comparison to the 4G wireless fronthaul/backhaul network. Similarly to the commercial zone, the power savings are the highest during the working hours. The results show that the 5G smart model achieves total power savings around 8.2% in comparison to the edge computing approach.

The impact of power reduction from NCC is negligible in comparison to the power reduction as consequence of the pBS placement. The reason lies on the fact that NCC reduces around 30% [17] of the power consumed in the MC phase, which is a small decrease when comparing to the power reduction of shutting down various mBS. Moreover, NCC highly impacts on the power consumed by the end user, but as explained in section 5, the power consumed in the UEs is assumed to be negligible in comparison to the power consumed in the BSs.

9.4. Sensitivity Benchmark

This subsection aims to understand how small changes in the network may affect the metrics evaluated in previous subsections. This helps to prevent a huge drop in network performance by proactively anticipating network changes. The smart model is used in the sensitivity benchmark since it performs better than the random model. In the evaluation, the metrics selected are: power consumed in the RAN, service efficiency, user distribution, pBS's user density, pBS density, and pBS coverage. The parameters changed in the sensitivity benchmark are user density, pBS range (or pBS transmission power), user mobility, and maximum pBS density. The mBS density remains the same for all the sensitivity benchmark. We decided to establish this assumption because of the time it would take to build another fixed antenna with a mBS to support peaks of traffic. While adding another drone can be done fast and easily, adding a new fixed antenna requires the assessment of the terrain and legal permissions to build new antennas. Moreover, this antenna would be hard to move if those peaks disappear. Therefore, we foresee

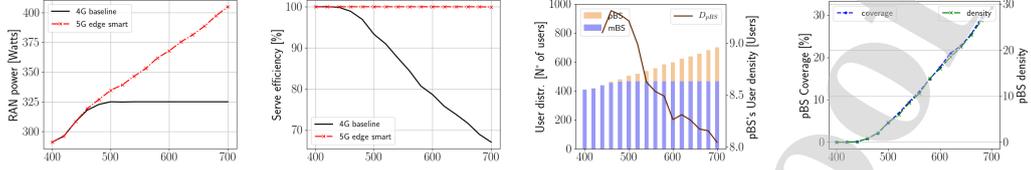
Description	Parameter	Value	Default
mBS density	N_{mBS}	4	4
User density	λ_{UE}	400..700	500
User mobility	m	0..1	0.5
pBS range	r_{pBS}	20..100	60
Maximum pBS density	N_{pBS}^{max}	10..30	30

Table 5: Parameters for simulation

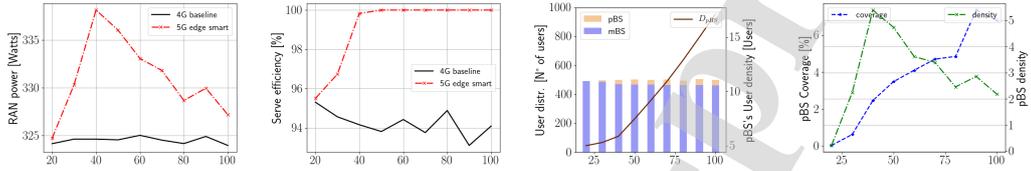
a scenario with enough mBS to cover the minimum traffic along the day and additional pBS to cover the peaks of traffic. Table 5 collects the values of the sensitivity parameters, their variation range, and their default value.

Taking the aforementioned assumption into consideration, in most of the cases, 4G would not be able to support the requested traffic. Consequently, it will have to reduce the bandwidth per user. This directly affects the serve efficiency, which is greatly reduced in the case of 4G. On the other hand, 5G would simply add more pBSs to support the extra traffic. This increases the power consumed in the RAN because a) there are more pBS in the field and b) users can enjoy more bandwidth and send data faster.

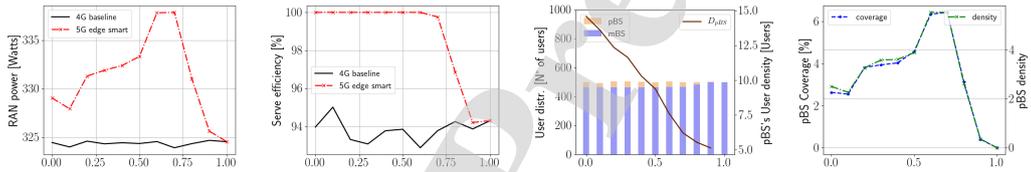
Fig. 9a shows the sensitivity benchmark for the power consumed in the RAN, user distribution, pBS's user density, pBS density, and pBS coverage for $\lambda_{UE} = \{300...700\}$. A linear correlation can be observed between the user density and the power consumed in the RAN. The 4G baseline reaches a limit on power consumption when the mBS infrastructure cannot support more traffic. Then, the power consumption in the 5G smart model increases and the power consumption in the 4G baseline stays the same. In other words, the users connect to the mBSs until the capacity limit is reached. Then the limit is reached, the 5G scenario can increase the capacity by adding more base stations, hence increasing the power consumption. On the other hand, the 4G baseline can only reduce the bandwidth sent to the users, either selecting to which user preserve the speed or reducing bandwidth uniformly if all users have the same priority. Consequently, the serve efficiency, which represents the percentage of resources received in comparison to the resources requested, is reduced. Further increases in user density increase the number of users connected to pBSs by increasing the pBS density and coverage, while



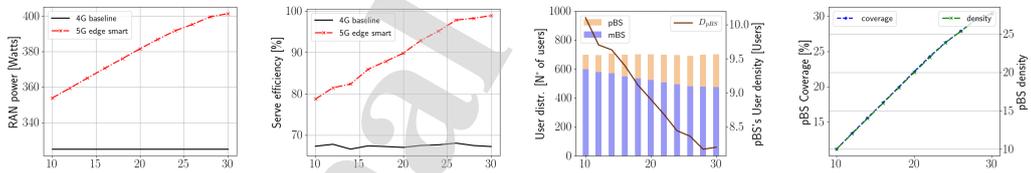
(a) Variations in user density: $\lambda_{UE} = \{300...700\}$.



(b) Variations in pBS range: $r_{pBS} = \{20...100\}$.



(c) Variations in user mobility: $m = \{0...1\}$.



(d) Variations in available pBS: $N_{pBS}^{max} = \{10...30\}$.

Figure 9: Sensitivity benchmark results of the power consumed in the RAN, the user distribution, the pBS's user density, the pBS density, and the pBS coverage.

reducing pBS's user density. The same behaviour of the reduction of pBS's user density is explained in sections 8 and 9.

Fig. 9b shows the sensitivity benchmark for the power consumed in the RAN, user distribution, pBS's user density, pBS density, and pBS coverage for $r_{pBS} = \{20...100\}$. The impact of the pBS range on the power consumed in the RAN can be divided into two well-defined steps clearly appreciated in Fig. 9b. In the first step, the pBS range is so small ($r_{pBS} \leq 40$) that the

power consumed of the new pBS does not cover for the user offload because user expectation inside the pBS is very small. Consequently, the pBS is not placed. Fig. 9b shows that pBSs with small ranges are not placed until the minimum user density per pBS reaches 5, which is the minimum pBS's user density in the scenario under evaluation. More pBS are able to cover at least 5 users with the increase in range ($r_{pBS} > 40$), which increases the number of pBSs in the network, and thus, the power consumed in the RAN. On the bright side, the increase in pBS density increases the bandwidth resources, thus increasing the capacity and improving service efficiency. The power consumption in the RAN reaches a maximum, which corresponds to the minimum value of r_{pBS} that optimizes the bandwidth resources. For bigger values of r_{pBS} , the bigger the range, the more the users under coverage. This reduces the number of pBS required and thus, the power consumed in the RAN.

Fig. 9c shows the sensitivity benchmark for the power consumed in the RAN, user distribution, pBS's user density, pBS density, and pBS coverage for $m = \{0..1\}$. The behavior of user mobility can be divided into two sections. In the first section ($m \leq 0.7$), there are enough static users to deploy the pBS accordingly. This maximizes the service efficiency. Furthermore, the higher the mobility rate, the smaller the pBS's user density (Fig. 9c), which increases the number of pBS required to cover the same number of users and thus, the power consumed in the RAN. In the second section ($m > 0.7$), the user's probability of being mobile is too high, which scatters the static users, which are the only users that can connect to a pBS. Occasionally, user sparsity can be very high that the controller cannot find a position to place a pBS that covers at least 5 users, and consequently, the bandwidth resources in the network are reduced. This also reduces service efficiency.

Fig. 9d shows the sensitivity benchmark for the power consumed in the RAN, user distribution, pBS's user density, pBS density, and pBS coverage for $N_{pBS}^{max} = \{10..30\}$. The idea of this sensitivity benchmark is to understand what would happen if there are not enough pBSs available in the system to cover the resources requested. Consequently, the network resources requested are increased by maximizing the user expectation ($\lambda_{UE} = 700$). The correlation between most of the evaluated metrics and the maximum pBS density is linear. The more pBSs available, the more pBSs used. This has a linear impact on the power consumed in the RAN and in the service efficiency. The pBS user density is inversely proportional to the N_{pBS}^{max} . This is easily

explained by observing the nature of the smart model since every new pBS placed will always have a smaller user density than the previous one. Consequently, the more pBSs, the smaller the pBS's user density. Fig. 9d also shows that the more pBSs placed, the more coverage, and the less pBS's user density. Consequently, it can be concluded that every new pBS covers the same extension, but the users in that region are sparser.

10. Conclusions

Network power savings is a critical issue in urban areas where the network infrastructure evolves to UDNs to cover the increase in traffic. NCC efficiently boosts the resources in the network with the use of MSCs, which are arrangements of UEs that offload the cellular traffic to underlay networks, however, the creation, placement, management, and destruction of MSCs is still an open research topic. In this article, we propose a methodology to efficiently activate, place, and deactivate MSCs in an urban area. To leverage daily network traffic changes, we divide the urban area into three different zones: residential, commercial, and manufacturing, and each representing a different network traffic distribution. The mobility nature of the MSC enables the inter-zone mobility. Consequently, the MSCs are moved to the positions where they are required the most at any time. As a result, the number of required MSCs is reduced. The investigation targets the reduction of power consumption in scenarios with two different RRM models (smart and random), which are validated with a Monte Carlo simulator.

The model has been validated against a Monte-Carlo simulator, achieving MSE values of 0.18 and 0.03 for the random and smart RRM schemes respectively. Moreover, the results show total power savings in the network for the smart model. The power savings are especially achieved in the RAN and the main contributor is the reduction of mBS density. NCC only impacts the performance in the pBSs and the UEs connected to them since they are the only BSs that can provide short-range communication and ad-hoc connectivity between users. The smart model achieves around 15 – 25% power savings in comparison to the 4G edge computing wireless fronthaul/backhaul and 6 – 15% power savings in comparison to the random model and the other two evaluated approaches that were taken from the related work. The smart model is six- to eleven-fold more efficient than the random model, depending on the hour of the day and the urban zone. In particular, the smart model achieves five times more user density per pBS than the random model. This

impacts directly on the power consumption. The proposed smart and random models require a UAV density of 90 and 832 UAVs/km² in the scenario under evaluation. Finally, a comprehensive sensitivity evaluation of different changes in the network is performed. The sensitivity results show that the user expectation and maximum pBS density impact linearly on the metrics studied. On the other hand, the pBS range and mobility of users present different behaviors. For example, the sensitivity evaluation shows that the range of the pBS (and thus, the transmission power) can be minimized to maximize the service efficiency in the RAN. In the case of the mobility factor, the sensitivity evaluation showed an upper bound in the mobility factor for maximizing the service efficiency in the RAN.

This paper provides a comprehensive study of MSCs placement for energy-efficient UDNs when NCC is used. However, evaluating this study in a real testbed is expected to raise new research challenges. In the future, we will continue this work through a testbed that validates the models proposed in this paper.

Acknowledgements

This research has been supported in part by European Union's H2020 research and innovation program under grant agreement H2020-MCSA-ITN-2016-SECRET 722424[51].

References

- [1] R. W. Heath and M. Kountouris, "Modeling heterogeneous network interference," in *2012 Information Theory and Applications Workshop*, 2012, pp. 17–22.
- [2] Cisco, "Cisco visual networking index: Forecast and methodology, 2017 to 2022," Cisco Technologies, Tech. Rep., Accessed: 2020.
- [3] Cisco Technologies, "Cisco visual networking index: Forecast and methodology, 2016 to 2021," Cisco Technologies, Tech. Rep., Accessed: 2020.
- [4] S. T. Baaran, G. K. Kurt, M. Uysal, and . Altunba, "A tutorial on network coded cooperation," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2970–2990.

- [5] S. Pandi, R. Torre, G. T. Nguyen, and F. H. P. Fitzek, "Massive Video Multicasting in Cellular Networks using Network Coded Cooperative Communication," in *2018 15th IEEE Annual Consumer Communications and Networking Conference (CCNC) (CCNC 2018)*, Las Vegas, USA, 2018.
- [6] J. Rodriguez, A. Radwan, C. Barbosa, F. H. Fitzek, R. A. Abd-Alhameed, J. M. Noras, S. M. Jones, I. Politis, P. Galiotos, and G. Schulte, "SECRETsecure network coding for reduced energy next generation mobile small cells: A european training network in wireless communications and networking for 5g," in *2017 Internet Technologies and Applications (ITA)*. IEEE, pp. 329–333, 00043.
- [7] X. Gelabert, P. Legg, and C. Qvarfordt, "Small cell densification requirements in high capacity future cellular networks," in *2013 IEEE International Conference on Communications Workshops (ICC)*, 2013, pp. 1112–1116.
- [8] P. Kela, X. Gelabert, J. Turkka, M. Costa, K. Heiska, K. Leppnen, and C. Qvarfordt, "Supporting mobility in 5g: A comparison between massive mimo and continuous ultra dense networks," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.
- [9] R. Bassoli, F. Granelli, S. T. Arzo, and M. D. Renzo, "Toward 5G cloud radio access network: An energy and latency perspective," *IEEE Access*, 2019.
- [10] W. Lu and M. Di Renzo, "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 179188. [Online]. Available: <https://doi.org/10.1145/2811587.2811597>
- [11] R. Ahlswede, Ning Cai, S. . R. Li, and R. W. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

- [12] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413–4430, 2006.
- [13] J. N. Laneman, *Chapter 1 COOPERATIVE DIVERSITY Models, Algorithms, and Architectures*.
- [14] F. Rossetto and M. Zorzi, "Mixing Network Coding and Cooperation for Reliable Wireless Communications," *IEEE Wireless Communications*, vol. 18, no. 1, pp. 15–21, 2011.
- [15] I. Leiva-Mayorga, R. Torre, S. Pandi, G. T. Nguyen, V. Plà Boscà, J. Martinez-Bauset, and F. H. P. Fitzek, "A Network-coded Cooperation Protocol for Efficient Massive Content Distribution," in *IEEE GLOBE-COM 2018 Conference Proceedings*, Technische Universität Dresden. Dubai, United Arab Emirates: IEEE, 2018.
- [16] I. Leiva-Mayorga, R. Torre, V. Plà Boscà, S. Pandi, G. T. Nguyen, J. Martinez-Bauset, and F. H. P. Fitzek, "Network-coded Cooperation and Multi-connectivity for Massive Content Delivery," *IEEE Access*, 2020.
- [17] R. Torre, I. Leiva-Mayorga, S. Pandi, H. Salah, G. Nguyen, and F. H. P. Fitzek, "Implementation of Network-Coded Cooperation for Energy Efficient Content Distribution in 5G Mobile Small Cells," *IEEE Access*, 2020.
- [18] D. Lopez-Perez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.
- [19] D. Liu, L. Wang, Y. Chen, M. ElKashlan, K. Wong, R. Schober, and L. Hanzo, "User association in 5g networks: A survey and an outlook," *IEEE Communications Surveys Tutorials*, vol. 18, no. 2, pp. 1018–1044, 2016.
- [20] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with rf energy harvesting: A contemporary survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 757–789, 2015.

- [21] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 550–560, 2012.
- [22] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.
- [23] X. Li and L. Xing, "Optimal deployment of drone base stations for cellular communication by network-based localization," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 7282–7287.
- [24] H. Huang, A. V. Savkin, M. Ding, and M. A. Kaafar, "Optimized deployment of autonomous drones to improve user experience in cellular networks," *Journal of Network and Computer Applications*, 2017.
- [25] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, "On the number and 3d placement of drone base stations in wireless cellular networks," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–6.
- [26] A. K. Mishra, J. L. Hellerstein, W. Cirne, and C. R. Das, "Towards characterizing cloud backend workloads: Insights from google compute clusters," *SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 4, p. 3441, Mar. 2010. [Online]. Available: <https://doi.org/10.1145/1773394.1773400>
- [27] Google Inc., "Open servers data for data processing," <http://googleresearch.blogspot.com/2010/01/google-cluster-data.html>, accessed: 2020-00-00.
- [28] Mapbox., "Manhattan Population Explorer," <http://manpopex.us/>, accessed: 2021-05-01.
- [29] Ericsson.com. Ericsson trials drone-mounted network. [Online]. Available: <https://www.ericsson.com/en/news/2020/12/ericsson-drone-mounted-network-could-support-emergency-response>
- [30] 3GPP, "3GPP TR 38.874 Study on Integrated Access and Backhaul (Rel. 16)," Tech. Rep., Accessed: 2020.

- [31] M. Di Renzo, W. Lu, and P. Guan, “The Intensity Matching Approach: A Tractable Stochastic Geometry Approximation to System-Level Analysis of Cellular Networks,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 5963–5983, 2016.
- [32] R. Torre, “Agile Mobile Edge Computing and Network-coded Cooperation in 5G,” Ph.D. dissertation, Technische Universitat Dresden, 2020.
- [33] Mitchell L. Moss and Carson Qing, “The Dynamic Population of Manhattan,” <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.392.6121>, accessed: 2021-05-01.
- [34] Metropolitan Transportation Authority (MTA), “Turnstile Data for Developers,” <http://web.mta.info/developers/turnstile.html/>, accessed: 2021-05-01.
- [35] New York Department of City Planning, “NYC Planning. Manhattan Zoning Maps,” <https://www1.nyc.gov/site/planning/zoning/index-map.page>, accessed: 2020-00-00.
- [36] Cisco Technologies, “Cisco visual networking index: Forecast and methodology, 2018 to 2023,” Cisco Technologies, Tech. Rep., Accessed: 2020.
- [37] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Gdor, G. Auer, and L. V. Der Perre, “Challenges and enabling technologies for energy aware mobile radio networks,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 66–72, 2010.
- [38] M. Lauridsen, L. Noël, T. B. Sørensen, and P. Mogensen, “An empirical LTE smartphone power model with a view to energy efficiency evolution,” *Intel[®] Technol. J.*, vol. 18, no. 1, pp. 172–193, 2014.
- [39] G. Auer, V. Giannini, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, C. Desset, and O. Blume, “Cellular Energy Efficiency Evaluation Framework,” in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–6.
- [40] G. P. Koudouridis, H. Lundqvist, H. Li, and X. Gelabert, “Energy efficiency of network-coding enabled mobile small cells,” *Computer*

- Communications*, vol. 121, pp. 50 – 58, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0140366417310368>
- [41] D. Sabella, A. de Domenico, E. Katranaras, M. A. Imran, M. di Girolamo, U. Salim, M. Lalam, K. Samdanis, and A. Maeder, “Energy Efficiency Benefits of RAN-as-a-Service Concept for a Cloud-Based 5G Mobile Network Infrastructure,” *IEEE Access*, vol. 2, pp. 1586–1597, 2014.
- [42] P. Ruiu, A. Bianco, C. Fiandrino, P. Giaccone, and D. Kliazovich, “Power comparison of cloud data center architectures,” in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.
- [43] T. Werthmann, H. Grob-Lipski, S. Scholz, and B. Haberland, “Task assignment strategies for pools of baseband computation units in 4G cellular networks,” in *2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 2714–2720.
- [44] B. Dai and W. Yu, “Energy Efficiency of Downlink Transmission Strategies for Cloud Radio Access Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 1037–1050, 2016.
- [45] 3GPP, “Study on channel model for frequencies from 0.5 to 100 GHz. TR 138 901 Release 14,” Tech. Rep., Accessed: 2020.
- [46] G. Bournaka, Y. Rahulamathavan, K. Cumanan, S. Lambbotharan, and F. Lazarakis, “Base station beamforming technique using multiple signal-to-interference plus noise ratio balancing criteria,” *IET Signal Processing*, vol. 9, no. 3, pp. 248–259, 2015.
- [47] C.-M. Tseng, C.-K. Chau, K. Elbassioni, and M. Khonji, “Autonomous Recharging and Flight Mission Planning for Battery-operated Autonomous Drones,” *arXiv e-prints*, p. arXiv:1703.10049, Mar. 2017.
- [48] Daniele Veneziano, “Distribution of the maximum of independent identically-distributed variables,” 2005, mIT. Probability and Statistic in Engineering Course.
- [49] 3GPP, “Physical layer procedures, TS36.213.” May 2016.

- [50] F. Jalali, K. Hinton, R. Ayre, T. Alpcan, and R. S. Tucker, "Fog computing may help to save energy in cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1728–1739, 2016.
- [51] J. Rodriguez *et al.*, "Secret-secure network coding for reduced energy next generation mobile small cells," *ITA Conference*, 2017.

Highlights

Power Efficient Mobile Small Cell Placement for Network-Coded Cooperation in UDNs

Roberto Torre, Muhammad Tayyab, George Koudouridis, Xavier Gelabert, Riccardo Bassoli, Frank H. P. Fitzek

- Mobile small cells will play a major role in the future Radio Access Network
- Network-coded Cooperation provides energy-efficient cellular content delivery
- Urban zoning city models help to exploit the mobility nature of mobile small cells
- Mobility in small cells enables on-demand deployment of network resources
- The power consumption in the network is significantly smaller than in 4G

Power Efficient Mobile Small Cell Placement for Network-coded Cooperation in UDNs

Roberto Torre^a, Muhammad Tayyab^b, George Koudouridis^c,
Xavier Gelabert^c, Riccardo Bassoli^a, Frank H.P. Fitzek^a

^a*Deutsche Telekom Chair of Communication Networks and Centre for Tactile Internet (CeTi) with Human-in-the-Loop, Georg-Schumann-Strae 11, Dresden, 01187, Saxony, Germany*

^b*Huawei Technologies Finland and Department of Communications and Networking, School of Electrical Engineering, Aalto University, Espoo, Finland*

^c*Huawei Technologies Sweden AB, Stockholm, Sweden*



ROBERTO TORRE is a Ph.D. student at the Deutsche Telekom Chair of Communication Networks at TU Dresden, working on the Europa SECRET project. He has been working with the chair since October 2016. He received his bachelor's degree from the University of Valladolid, Spain in 2015. He continues his studies in the field of telecommunications and he obtained his master's degree in 2017 from the University of Valladolid, Spain. During his studies, he developed deep interests in neural networks, GPU computing, network coding, and mesh networks. His research interests are Mobile Edge Computing, Software Defined Networking, and Network Function Virtualization.

Email address: roberto.torre@tu-dresden.de (Frank H.P. Fitzek)



MUHAMMAD TAYYAB received his M.Sc. degree in electrical engineering from King Fahd University of Petroleum and Minerals, Saudi Arabia, in 2017. He is currently pursuing a Ph.D. degree in electrical engineering with Aalto University, Finland. He has more than two years, from 2013 to 2015, of professional experience as an RF Planning and Optimization Executive at Wi-Tribe Pakistan Ltd. (Ooredoo Group). He has been a Researcher with the Helsinki Research Center, Huawei Technologies Finland Oy, since February 2018. His current research interest includes energy-efficient mobility for small-cell overlaid cellular networks.



GEORGE P. KOUDOURIDIS received a. degree in computer sciences from the Department of Computer and Systems Sciences, Stockholm University (SU), Stockholm, Sweden, in 1995, and a Ph.D. degree in electrical engineering on telecommunications from the Royal Institute of Technology (KTH), Stockholm, in 2016. He is currently a Principal Researcher with the Radio Access Network Systems Laboratory, Huawei Technologies Sweden, Stockholm Research Centre. Prior to joining Huawei in 2008, he was with Telia Research and TeliaSonera Mobile Networks Research and Development. His research interests include radio resource management, spectrum allocation, self-organized networks, artificial intelligence and machine learning based optimization in wireless networks.



XAVIER GELABERT received a double M.S. degree in electrical and telecommunications engineering from the Royal Institute of Technology (KTH) and the Universitat Politècnica de Catalunya (UPC) in 2004. He also received a Ph.D. degree from the UPC in 2010. Since 2012 he has been a senior research engineer at Huawei Stockholm Research Centre. He has more than 15 years of experience in wireless research. His interests include radio and spectrum resource management, self-organized networks, and more recently baseband systems implementation and design. He has been actively contributing to 3GPP NR standards in RAN1 and RAN2.



Riccardo Bassoli is a senior researcher with the Deutsche Telekom Chair of Communication Networks at the Faculty of Electrical and Computer Engineering, Technische Universität Dresden. He received his B.Sc. and M.Sc. degrees in Telecommunications Engineering from University of Modena and Reggio Emilia in 2008 and 2010 respectively. He received his Ph.D. degree at University of Surrey, in 2016. Between 2011 and 2015, he was also a Marie Curie Early Stage Researcher at Instituto de Telecomunicações (Portugal) and a visiting researcher at Airbus Defence and Space (France). Between 2016 and 2019, he was a postdoctoral researcher at the University of Trento.



FRANK H. P. FITZEK is a Professor and head of the Deutsche Telekom Chair of Communication Networks at TU Dresden coordinating the 5G Lab Germany. He is the spokesman of the DFG Cluster of Excellence CeTI. He received his diploma degree in electrical engineering from the University of Technology Rheinisch-Westflische Technische Hochschule Aachen, Germany, in 1997 and his Ph.D. in Electrical Engineering from the Technical University Berlin, Germany in 2002. In 2003 he joined Aalborg University as an associate professor and later became a professor.

Author Statement

We hereby certify that we have authored this paper independently and without undue assistance from third parties. No other than the resources and references indicated in this paper have been used. We have marked both literal and accordingly adopted quotations as such. There were no additional persons involved in the intellectual preparation of the present paper.

The following authors were involved in the following activities.

Roberto Torre: Conceptualization, methodology, software, analysis, validation, edition.

Muhammad Tayyab: methodology, contextualization, conceptualization support.

George Koudouridis: Conceptualization, contextualization, methodology, revision.

Xavier Gelabert: Conceptualization, contextualization, methodology, revision.

Riccardo Bassoli: Conceptualization, contextualization, methodology, revision.

Frank H.P. Fitzek: Revision, resources, funding

Conflict of Interest

There are not conflict of interest

Journal Pre-proof