Modeling Mobile Edge Computing Deployments for Low Latency Multimedia Services

Jorge Martín-Pérez*,†, Luca Cominardi†, Carlos J. Bernardos†, Antonio de la Oliva†, Arturo Azcorra†‡
†Universidad Carlos III de Madrid, Spain
‡IMDEA Networks Institute, Spain

Abstract—Multi-access Edge Computing (MEC) technologies bring important improvements in terms of network bandwidth, latency and use of context information, critical for services like multimedia streaming, augmented and virtual reality. In future deployments, operators will need to decide how many MEC Points of Presence (PoPs) are needed and where to deploy them, also considering the number of base stations needed to support the expected traffic. This article presents an application of inhomogeneous Poisson point processes with hard-core repulsion to model feasible MEC infrastructure deployments. With the presented methodology a mobile network operator knows where to locate the MEC PoPs and associated base stations to support a given set of services. We evaluate our model with simulations in realistic scenarios, namely Madrid city center, an industrial area, and a rural area.


I. INTRODUCTION

The next generation of mobile networks, the so-called 5G, comes with the promise of new and enhanced services, through the introduction of an improved radio interface, plus a network core that allows to dynamically deploy services closer to the location of the users. 5G networks will need to accommodate on top of the same physical infrastructure multiple kinds of services with very distinct requirements, spanning from ultra-low latency to high bandwidth and high reliability. These services are grouped in three main categories by 3GPP: enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low Latency Communications (URLLC), and Massive Internet of Things (MIoT) [1].

Among the several use cases that may be supported by 5G are Augmented Reality (AR) and Virtual Reality (VR), which can be included into eMBB and URLLC categories. In particular, AR/VR impose a Motion to Photon (MTP) latency that does not exceed 20 ms, requiring a network Round Trip Time (RTT) below 2 ms [2]. Moreover, a response within 1 ms is desired in case of visual-haptic interaction [3]. Although the new 5G radio interface promises ultra low latency enhancements, to truly fulfill the VR/AR ultra-low latency requirements it is also necessary to reduce the communication distance by bringing the multimedia applications close to the end users. This is achieved by Multi-access (Mobile) Edge Computing (MEC) [4] [5].

MEC is a key enabler for 5G technology and its main principle is to host computation and storage at edge hosts, close to end users. Typically, these edge hosts are highly distributed in the network, located close to the radio access network nodes (e.g., gNodeBs in 5G). As a result, MEC enables two types of services: (i) low-latency services, requiring a very low and bounded delay between the end user device and the server hosting the application; and, (ii) context-aware services, which need to access end-user contexts, such as the user channel quality conditions, in order to adapt the delivered service. MEC is being standardized within ETSI, via the MEC ISG group. A simplified version of the MEC architecture defined by ETSI is shown in Fig. 1. The main components of the architecture include: the Mobile Edge (ME) host, the ME application and the MEO (ME orchestrator). The MEC host is the key element. It provides the environment to run ME applications, while it interacts with the mobile network entities, via the Mobile Edge Platform (MEP), to provide ME services and deliver mobile traffic to MEC applications.

ME hosts are expected to be deployed by mobile operators in their 5G network infrastructure. To enable the pervasive service offering of AR/VR multimedia services, it is hence necessary to study how a MEC deployment should look like to support URLLC. Specifically, it is important to understand what are the suitable locations of future MEC Points of Presence (PoP) within the mobile network infrastructure. As a result, in this article we develop a novel model for MEC

*Corresponding author: jmartinp@it.uc3m.es
Parts of this paper have been published in the Proceedings of the IEEE BMSB 2018, Valencia, Spain.
Work partially funded by the EU H2020 5G-CORAL Project (grant no. 761586) and by the EU H2020 5G-TRANSFORMER Project (grant no. 761536).
scenario deployments, analyzing how many MEC Points of Presence (PoPs) are needed to meet a given set of requirements.

This work has been done within the framework of the 5G-TRANSFORMER project. The 5G-TRANSFORMER main goal is to support vertical industries (particularly focusing on low latency) through flexible slicing and federation of resources across multiple domains. This article is structured as follows: Section II introduces related work, Section III presents in detail our model, which is then applied to a realistic scenario in Section IV and validated in Section V. Finally, Section VI concludes this article with some final remarks and hints for future work.

II. RELATED WORK

Existing work in the literature, such as [6], studies feasible network infrastructure deployments using Point Processes (PPs) that randomly scatter points on some space (e.g., a line, a Cartesian plane, etc.). In particular [6] uses Poisson PPs (PPPs), a family of PPs used to generate the location of base stations (BSs) that are distributed following Poisson counting processes. PPPs can impose a minimum distance between BSs to increase coverage area and reduce the interference (see hard-core PPPs in [7]), and control if every region in the space has the same or different probability to host a BS (see homogeneous and inhomogeneous PPPs, respectively, in [7]).

Different types of PPPs are used in the State of the Art (SoA). For instance, the authors of [8] use Neyman-Scott [9] PPPs to generate small base stations (BSs) clustered around macro BSs with the objective of modeling the coverage and interference in heterogeneous networks. Similarly, [10] shows that homogeneous PPPs can be used in realistic deployment scenarios to reduce interference and increase the coverage area by properly configuring the distance between the BS sites and the intensity parameter.

Works like [6] and [11] analyze potential infrastructure deployments with the help of homogeneous PPPs. The former focuses on studying the cost of a Cloud Radio Access Network (Cloud-RAN) infrastructure using PPPs, while the latter analyzes the throughput and coverage of end users, using PPPs to determine hotspot locations in heterogeneous cellular networks.

Regarding future MEC deployments in real scenarios, [12] studies the existing BS deployment in several USA locations (including highways and rural areas). Likewise, [13] characterizes AR/VR deployments in stadium scenarios in the perspective of forthcoming Tokyo 2020 Olympics. The authors conclude that such use case requires a single media room dedicated to the video processing and proposes two distinct BS deployments: one powerful BS in the stadium (e.g., on the roof) or multiple small BSs (e.g., located close to the stadium vomitoria$^1$).

The work presented in this article studies the deployment of MEC PoPs in real urban, industrial and rural scenarios to satisfy the 5G slices requirements, including the most stringent latency expected by the 5G AR/VR applications (as done in [13]).

Unlike [12], this paper accounts for 5G New Radio (NR) technologies rather than legacy radio technologies. We generate feasible locations for gNodeBs, and derive deployments of MEC PoPs guaranteeing low latency requirements that deployments in [12] do not ensure. We use inhomogeneous Matérn II PPPs (see Proposition III.2 in Sec. III-A for more details) to obtain feasible locations for the BSs. The inhomogeneity of these processes improves the Matérn PPPs in the SoA by locating the BSs based on the density of population. That is, more BSs are generated where the population density$^2$ is higher. In other words, the inhomogeneity allows to concentrate the BSs where they are really needed and to minimize the generation in those areas with little traffic demand.

The generated BS locations are used next to derive the required number of PoPs and their potential location.

III. MODEL

This section presents the model used to both generate the BSs of future MEC deployments, and to determine the locations of the MEC PoPs to which the BSs are assigned. In our model a BS is the first connection point in the RAN accessed by the User Equipment (UE), for example a 2G BS, a 3G NodeB, a 4G eNodeB, or a 5G gNodeB. In Sec. III-A we extend Matérn PPPs to introduce inhomogeneity and generate BSs based on the population density. The derived processes, namely inhomogeneous Matérn I and II PPPs, impose a minimum distance between BSs to obtain a better coverage. Moreover, we show that Matérn II PPPs are more suitable for BSs generation because of their relaxed thinning procedure. Sec. III-B presents the algorithm used to generate MEC PoPs that can satisfy imposed latency constraints among all the BSs that are assigned to them.

A. Macro cells generation

With the model we aim to generate more BSs in regions where the population is higher. Here we refer to a region $R$ as a subset of a complete separable metric space $\mathcal{R}$, for example a map of Spain. Hence $R$ can represent a city like Madrid and the model locates BSs in city areas where there are more people. To achieve this we consider that every region $R$ contains population circles $C_i \subset R$ defined by a center $c_i$ and a revolution function $f_i(x)$ (with $i \in [1, R_C] \cap \mathbb{N}$, and $R_C$ the number of population circles defined inside the region $R$).

The revolution function $f_i(x)$ with $x \in C_i$ determines where it is more likely to have a person in $C_i$. This function leads to a surface of revolution defined at $C_i$, that can be expressed as $f_i(||x - c_i||_d)$, and which height expresses the amount of people at a given location $x \in R$. Therefore, the presence of people at $x \in R$ is expressed as:

$$G(x) := \sum_i f_i(||x - c_i||_d), \quad \forall C_i \subset R, \forall d \in \mathbb{N}$$

where $G(x)$ is referred as gentrification in the following paragraphs.

$^1$A vomitorium is a passage situated below a tier of seats in a stadium.

$^2$In our scenarios we refer to human population. Nevertheless, inhomogeneous Matérn PPPs can be applied to other types of populations.
Figure 2 illustrates an example of a region \( R \) with two population circles (i.e., \( C_1, C_2 \)) corresponding to two distinct areas. Moreover, the revolution functions \( f_1, f_2 \) considered in this example are Gaussian-like surfaces of revolution that can be multi lobed as the case of \( f_2 \), and the gentrification function \( G(x) \) is just the dashed line merging both of them.

If an inhomogeneous PPP uses the gentrification function \( G(x) \), the process generates BSs in concordance with the population, but it is still necessary to know how many we have to generate. Based on [14] we grid the region \( R \) in cells to satisfy the imposed number of BSs per \( \text{km}^2 \) in future 5G deployments, and we take \( R \) as a two dimensional space where the region to be gridded \( R \subset \mathbb{R} \) (for example Madrid) is expressed as a rectangle \( R = [x_{1,i}, x_{1,j}] \times [x_{2,i}, x_{2,j}] \) divided into cells \( R_i \subset R \) of width \( x_{1,i} \) and height \( x_{2,i} \). The resulting grid has \( w_i \) rows and \( u_i \) columns that are completely determined by the cell index \( i \) (which increases from left to right, and upside down as shown in Fig. 3):

\[
w_i := \left\lfloor \frac{i \cdot x_{1,i}}{u_i} \right\rfloor, \quad u_i = \left\lfloor \frac{i \cdot x_{1,i}}{u_i} \right\rfloor \mod u_n \tag{2}
\]

\[
u_n = \left\lfloor \frac{x_{1,i} - x_{1,j}}{x_{1,i}} \right\rfloor \tag{3}
\]

with \( u_n \) being the number of columns of the gridded region \( R \). Finally, we set each cell \( R_i \) as the product of two intervals \( R_i = [x_{1,i}, x_{1,j}] \times [x_{2,i}, x_{2,j}] \) whose limits are:

\[
x_{1,i} = x_{1,i} + x_{1,i}u_i \tag{4}
\]

\[
x_{1,i} = \min \{ x_{1,i}, x_{1,i} + (1 + u_i)x_{1,i} \} \tag{5}
\]

\[
x_{2,i} = \max \{ x_{2,i}, x_{2,i} + (1 + u_i)x_{2,i} \} \tag{6}
\]

\[
x_{2,i} = x_{2,i} - x_{2,i}u_i \tag{7}
\]

Fig. 3 illustrates on the left-hand side (lhs) the limiting coordinates of a region \( R \) that is gridded into cells \( R_i \) according to the limits above, i.e., (4)-(7).

In every single cell an inhomogeneous PPP generates a specific average number of BSs using an intensity function defined as \( \lambda(x) = k \cdot G(x) \) (where \( k \in \mathbb{N} \) is a constant), so BSs are located with higher probability where there are more people. But since we want to have the BSs as sparse as possible, it is necessary to impose a minimum distance between them. This work leverages the Matérn hard-core processes [15] to model the BSs’ generation. The first process under consideration for our model is the Matérn I process.

**Definition III.1.** Matérn I point process: is the point process obtained after applying a thinning with index function:

\[
I_1(x, X) := \begin{cases} 
0 & \text{if } N(B(x, r)) > 1 \\
1 & \text{if } N(B(x, r)) = 1 
\end{cases} \tag{8}
\]

to a stationary PPP \( X \), where \( N(B(x, r)) \) denotes the number of points of the point process \( X \) falling in the ball centered at \( x \) with radius \( r \).

In other words, a point \( x \in X \) is removed if it has a neighbor \( x' \in X \) with distance \( ||x - x'||_d < r \). This property suits the random generation of the BSs’ locations because only one BS can be in a neighborhood. However, these point processes are stationary (see [7]), and therefore homogeneous, before the thinning (see Definition III.1), but this model uses inhomogeneous PPPs to generate BSs (based on \( G(x) \)) at each cell \( R_i \subset R \). Thus we modify the original definition of a Matérn I point process to use an inhomogeneous PPP before the thinning. We call these processes inhomogeneous Matérn I PPPs.

After inhomogeneity is introduced, we need to know an expression for the average number of points \( \mathbb{E}[N(C)] \) that can appear in a certain set \( C \subset R \). The reason, as shown in Proposition III.1, is that this expression is in terms of the repulsion radius \( r \) and the intensity function \( \lambda(x) \), and we must select these parameters accordingly to generate the desired average number of BSs in a cell \( R_i \). Since there is no expression in the literature for the average number of points of a “inhomogeneous Matérn I PPPs”, we have obtained it with the help of the Campbell-Mecke formula [16].

**Proposition III.1.** Given an inhomogeneous PPP \( X \) with intensity function \( \lambda(x) \), and the thinning function \( I_1 \), the resulting thinned point process, called inhomogeneous Matérn I PPP, has the following average number of points at \( C \):

\[
\mathbb{E}[N(C)] := \int_C e^{-\int B(x, r)} \lambda(u) du \lambda(x) dx \tag{9}
\]

where \( r \) is the thinning radius of \( I_1 \).

**Proof.** First we define the auxiliary function \( g \):

\[
g : \mathcal{R} \times \Omega \to \{0, 1\} \tag{10}
\]

\[
(x, A) \mapsto \mathbb{1}_{C}(x) \mathbb{1}(\text{dist}(x, X \setminus A) > r) \tag{11}
\]

where \( \Omega \) is the space of events (see [7]). We can then rewrite the average number of points at \( C \) as:

\[
\mathbb{E}[N(C)] := \mathbb{E} \sum_{x \in X} g(x, X) \tag{12}
\]
Next, we use the Campbell-Mecke formula to obtain:

\[
\mathbb{E} \sum_{x \in X} g(x, X) := \int_{\mathbb{R}} \mathbb{E}^x [g(x, X)] \lambda(x) \, dx
\]

(13)

\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, X) \mathbb{P}^x(A) \, dA \lambda(x) \, dx
\]

(14)

\[
= \int_{\mathbb{C}} \int_{\Omega} \mathbf{1}(\text{dist}(x, X \setminus A) > r) \, dA \lambda(x) \, dx
\]

(15)

\[
= \int_{\mathbb{C}} \mathbb{P}^x(\text{dist}(x, X \setminus A) > r) \lambda(x) \, dx
\]

(16)

\[
= \int_{\mathbb{C}} \mathbb{P}(N(B(x, r)) = 0) \lambda(x) \, dx
\]

(17)

where \(\mathbb{P}^x\) denotes the Palm probability. Finally, we apply the capacity functional (see [7]) of a PPP to obtain the stated equality.

One drawback of these “inhomogeneous Matérn I” processes is their very restrictive dependent thinning procedure, which might reach the case where all the points are removed in certain neighborhoods. To overcome such limitation, our model considers a second type of processes, known as Matérn II point processes, that rely on marked point processes (see [7]). Matérn II processes assign a mark to every point generated so as to allow the dependent thinning processes (see [7]) distinguish which point is retained in a neighborhood.

Definition III.2. Matérn II point process: is the point process obtained after applying a thinning with index function:

\[
I_2(x, m, X, M_X) := \begin{cases} 
1 & \text{if } m = \min_{m' \in M_X} \{x', m'\} : x' \in B(x, r) \\
0 & \text{otherwise}
\end{cases}
\]

(18)

to a stationary marked PPP \(X\), where \(M_X\) denote the marks associated to the point process \(X\).

In other words, among all the points falling in the ball of radius \(r\), only the one with the lowest mark \(m\) survives. These kind of processes present the advantage that even when the intensity function takes high values, at least one point remains in every neighborhood. In our model this means that in a certain neighborhood there is no more than one BS.

If rather than using a stationary PPP before applying the dependent thinning \(I_2\), we use an inhomogeneous PPP (something novel in the SoA); then it is possible that the retained BS in a neighborhood is the one with higher \(\lambda(x)\) by choosing the mark \((x, m)\) as \(m \sim \frac{1}{\lambda(x)}\). But still is missing how we can control that a correct number of BSs is generated at each region \(R_i \subset R\) based on the repulsion radius \(r\) and the intensity function \(\lambda(x)\). Thus we proceed as with the “inhomogeneous Matérn I” PPs to obtain an expression for the average number of points (something novel in the SoA).

Proposition III.2. Given an inhomogeneous marked PPP \(X\) with intensity function \(\lambda\), the thinning function \(I_2\), and marks \(m \sim \frac{1}{\lambda(x)}\), the resulting thinned point process, called inhomogeneous Matérn II PP, has the following average number of points at \(C\):

\[
\mathbb{E}[N(C)] := \int_C e^{-\int_{B(x, r)} \mathbf{1}(\lambda(u) > \lambda(x)) \lambda(u) \, du} \lambda(x) \, dx
\]

(19)

where \(r\) is the thinning radius of \(I_2\).

Proof. As in the proof of Proposition III.1, we proceed defining the function \(g\):

\[
g : \mathbb{R} \times \Omega \to \{0, 1\}
\]

(20)

\[
(x, A) \mapsto \mathbb{I}_C(x) \mathbb{I} \left( \lambda(x) = \max_{x' \in X \cap B(x, r)} \{\lambda(x')\} \right)
\]

(21)

Then, the average number of points in a subset \(C\) can be expressed as in Eq. (12). Next, we apply the Campbell-Mecke
formula as in the previous proof and we obtain the average number of points as:

$$
\mathbb{E} \sum_{x \in X} g(x, X) := \int_C \mathbb{P} \left( \lambda(x) = \max_{x' \in X \cap B(x, r)} \lambda(x) \right) \, dx = 
$$

(22)

$$
= \int_C \mathbb{P} \left( N \left( B_{>\lambda(x)}(x, r) = 0 \right) \lambda(x) \, dx = 
$$

(23)

$$
= \int_C e^{-f_{B(x, r)}} \lambda(x) |(\lambda(u) > \lambda(x))| du \lambda(x) \, dx 
$$

(24)

where $B_{>\lambda(x)}(x, r) = \{ u : x \in B(x, r) \wedge \lambda(u) > \lambda(x) \}$.

The right-hand side (rhs) of Fig. 3 depicts how to obtain BSs’ locations using an inhomogeneous Matérn II PP (whose average number of points is derived in Proposition III.2). First, it generates the gray crosses using an inhomogeneous PPP with intensity function $\lambda(x)$ (note that more points are generated on the upper right corner because of the direction of $\nabla \lambda(x)$). Then, a $I_2$ thinning is applied using repulsion radius $r$ with marks $m = \frac{1}{\lambda(x)}$, so only the gray cross with the most top-right coordinate falling within a ball of radius $r$ survive. This surviving cross is hence illustrated as a cell tower in Fig. 3.

As both Propositions III.1 and III.2 state, the average number of points of “inhomogeneous Matérn I and II” PPs depend on the repulsion radius $r$. Then for a cell $R_i \subset R$, we need to know which $r$ allows the generation of the required number of BSs. To generate $N(R_i)$ points we set the cell repulsion radius as:

$$
r := \frac{2 \sqrt{x_{1,1} \cdot x_{2,2}}}{\sqrt{N(R_i)}}
$$

(25)

which grids $R_i$ in cells that can contain the repulsion balls $B(x, r)$ of the Matérn PPs. These new cells are squares of size $2r$ and area $4r^2 > \pi r^2 = |B(x, r)|$, and they divide the parent cell $R_i$ in $\sqrt{N(R_i)}^2$ smaller cells that can host a whole macro cell repulsion area.

B. MEC PoPs’ deployment

We assign the traffic of every BS generated in Sec. III-A to a MEC PoP that is deployed somewhere within the operators’ network infrastructure, and at a specific geographic location (i.e., to minimize RTT). We further divide the distance between the MEC PoP at location $m$, and a BS at location $x$, the higher the propagation delay $l(\cdot)$ of the link connecting them. The other two contributions for the packet Round Trip Time (RTT) are the radio transmission delay $t_r$ between a BS and the final user, and the packet processing delay $p(\cdot)$:

$$
RTT := 2l(\|x - m\|_d) + 2p(M) + t_r
$$

(26)

with $p(M)$ denoting the processing delay introduced by the network hops to be traversed to reach the network ring $M$. Hence, our model envisions the operator infrastructure as a hierarchy of network rings $\mathcal{M}$, in which traffic traverses more hops to reach network rings that are higher in the hierarchy (see Fig. 4). That is, taking $<_r$ as a relationship expressing which network ring is higher in the network hierarchy $\mathcal{M}$, if $M_a, M_b \in \mathcal{M}$ and $M_a < M_b$, we can say that a MEC PoP deployed at $M_a$ is reached in less hops than one at $M_b$. Thus, $p(M_a) < p(M_b)$, and network ring $M_a$ aggregates less traffic than $M_b$.

So if we place a MEC PoP at location $m$ and assign it to network ring $M$, after fixing the maximum RTT and radio technology, from Eq. (26) we know the maximum distance to those BSs whose traffic is assigned to the new MEC PoP

$$
\|x - m\|_d \leq \min \left\{ \frac{RTT - 2p(M) - t_r}{2} \right\} = m_M
$$

(27)

Algorithm 1 determines the MEC PoPs’ deployment in three stages:

1) **Initialization**: this first stage (lines 1-4) creates the set of MEC PoPs locations, the set of their network rings, the set of BSs and one matrix of locations $x \in R$ per network ring. Each entry $x$ in the matrix $\text{matrices}[M]$ represents the number of BSs that can be assigned to a MEC PoP deployed at a given location $x$ and associated to the ring $M \in \mathcal{M}$.

2) **Candidates search**: this second stage (lines 5-10) determines how many BSs can be assigned to a MEC PoP depending on its location and associated network ring $M \in \mathcal{M}$. For every $M$ it loops through each BS and increases by one those location entries of $\text{matrices}[M]$ satisfying that if a MEC PoP is deployed there, it could satisfy the RTT and hence, the BS could be assigned to that MEC PoP. If $\text{matrices}[M_0][x_0] = 4$ after this stage, it would mean that a MEC PoP associated to $M_0$ and deployed at $x_0$ would have 4 BSs assigned to itself.

3) **MEC PoPs selection**: this third and last stage (lines 11-35) is the main loop of Algorithm 1, and each iteration decides a new MEC PoP location and network ring. It comprises two phases:

a) **MEC PoP location**: this phase (lines 15-24) obtains the best location where a MEC PoP can be located. It iterates through all the possible network rings in $\mathcal{M}$ searching for the location where the maximum amount of BSs can be assigned to a MEC PoP. In case of having multiple locations assigning the same amount of BSs at different rings, let’s call them $M_a, M_b$, then the algorithm selects the location related to the ring with the minimum propagation delay, i.e., $\min \{ p(M_a), p(M_b) \}$. b) **Assignment update**: the MEC PoP obtained in the previous step handles the traffic of up to maximum number of BSs $\text{ringMaxBSs}(\cdot)$, depending on the network ring where it is associated. Taking into account that consideration, this phase (lines 28-34) iterates through every BS assigned to the MEC PoP and updates every $\text{matrices}[M]$ to reflect the assignment by decreasing in one unit the neighborhood of each BS. That is, if the new MEC PoP has an assigned BS at location $x$, the neighboring locations of $x$ at every ring, i.e., $\text{matrices}[M][B(x, m_M)]$, must be decreased by one. Thus if another MEC PoP is latter located inside $B(x, m_M)$, it knows that it will cover one less BS. Finally all the assigned BSs are removed from the unassigned list.
Hence, Algorithm 1 iterates through every MEC PoP candidate location \( m \), and chooses the one maximizing the number of BSs falling inside the ball \( B(m, m_M) \). This process minimizes the number of MEC PoPs, and is done for every possible network ring \( M \) to find the best \((m, M)\) combination.

**Data:** BSs, \( R \), RTT  
**Result:** MECPoPLocations, MECPoPRings

```
1 matrices = array(int.matrix(R), length = |M|);  
2 unassigned = Set(BSs);  
3 MECPoPLocations = Set();  
4 MECPoPRings = Set();  
5 foreach  
6   | in BSs do  
7     | m_M = l^{-1}\left(\frac{R_{TT} - 2p(ring) - t_r}{2}\right);  
8     | matrices[M][B(x, m_M)] += 1;  
9   end  
10 while not empty unassigned do  
11   | covBSs = -1;  
12   | MECPoP = NULL;  
13   | ring = NULL;  
14   | foreach  
15     | M in M do  
16        | maxCov = max_{x'}{\{matrices[M][x']\}};  
17        | moreBSsCovered = maxCov < covBSs;  
18        | eQ = (maxCov = covBSs \land p(M) < p(ring));  
19        | if moreBSsCovered OR eQ then  
20           | covBSs = maxCov;  
21           | MECPoP = x : matrices[M][x] = maxCov;  
22           | ring = M;  
23        end  
24   | MECPoPLocations.add(MECPoP);  
25   | MECPoPRings.add(ring);  
26   | ringMaxDis = l^{-1}\left(\frac{R_{TT} - 2p(ring) - t_r}{2}\right);  
27   | assignBSs = BSs \cap B(MECPoP, ringMaxDis);  
28   | foreach  
29     | x in assignBSs subset(ringMaxDis) do  
30         | foreach  
31           | M in M do  
32             | matrices[M][B(x, m_M)] -=1;  
33         end  
34   | unassigned.pop(x);  
35 end  
```

**Algorithm 1:** MEC PoPs placement.

### IV. Applying the model to a realistic scenario

This section provides an example of applying the model described in Sec. III. Specifically, Sec. IV-A characterize the network traffic and infrastructure to derive realistic values for the average number of BSs/km² and network RTT. Then, Section IV-B characterizes three deployment areas in Spain, namely Madrid city center (104 km²), Cobo Calleja industrial estate (8 km²), and Hoces del Cabriel valley (2193 km²); where the proposed model is applied.

### A. Characterization of network traffic and infrastructure

According to the Next Generation Mobile Network Alliance (NGMN) [17], 5G services are expected to be provided via ad-hoc network slices over the same physical infrastructure. In order to enable the desired traffic treatment in the network infrastructure, the 3rd Generation Private Partnership (3GPP) has defined a set of flows with the corresponding traffic requirements for the eMBB and URLLC slices [18], while NGMN in [19] defined the traffic requirements for the MloT slice. eMBB services are characterized by high bandwidth and span from classical mobile traffic (e.g., mobile terminals), to broadcast-like services (e.g., IPTV), high-speed vehicles (e.g., in-vehicle infotainment), indoor hotspot (e.g., fiber-like access), and dense urban (e.g., crows in a stadium, square). On the contrary, URLLC services are characterized by low latency and span from discrete automation (i.e., remote-controlled robots), to intelligent transport systems (e.g., autonomous cars), and tactile interaction (e.g., augmented reality). Finally, MloT services are characterized by a high number of intermittent and low-power communications (e.g., sensors). Table I reports a selected number of the above traffic requirements as reported in [18], [19].

Based on the different slices and traffic flows introduced above, the authors in [14] first identify three reference deployment scenarios, namely urban, industrial, and rural. Next, they characterize the average number of BSs/km² for each scenario. Specifically, they report for the urban scenario an average number of 72 BS/km² in case of supporting the indoor hotspot traffic flow, which is characteristic of business districts and office areas that require 4 BSs per building floor. In residential/commercial areas instead, the average number of BS/km² is 12 in urban scenarios. Similarly, 12 BSs are also required in the industrial scenario to satisfy the traffic demand of 1 km². Finally, the rural scenario considers a 4-lane road (e.g., highway) supporting the intelligent transport system flow (e.g., V2X) and requires 1 BS per kilometer of road.

The next step is to characterize the network infrastructure. To that end, we leverage the reference network infrastructure illustrated in [14] and based on [20]. The network infrastructure comprises three segments: (i) access, (ii) aggregation, and (iii) core. The access comprises 6 BSs for each node M1 connected via a point-to-point link, and 6 nodes M1 connected in a ring topology. Thus, each access ring hence connects a total of 36 BSs. It is worth highlighting that from a network

<table>
<thead>
<tr>
<th>SLICE</th>
<th>FLOW</th>
<th>REQUIREMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor Hotspot</td>
<td>Up to 1Gbps/user</td>
<td></td>
</tr>
<tr>
<td>eMBB</td>
<td>Broadcast services</td>
<td>Up to 200 Mbps/TV channel</td>
</tr>
<tr>
<td>High-speed</td>
<td>Vehicle</td>
<td>Up to 100 Gbps/Km²</td>
</tr>
<tr>
<td>URLLC</td>
<td>Discrete automation</td>
<td>Maximum jitter of 1 µs</td>
</tr>
<tr>
<td>MloT</td>
<td>Sensors data</td>
<td>Up to 200 Mbps/Km²</td>
</tr>
</tbody>
</table>

1 Note: end-to-end maximum delay as defined in [18].
In case of collocating the MEC PoP with M1 nodes, each of which serves 4 access rings. In turn, each aggregation ring is served by two M3 nodes for redundancy reasons, while each M3 node provides gateway capabilities to 2 aggregation rings. Finally, M4 nodes are connected to the core ring and serve as gateway to M3 nodes. According to the reference network infrastructure, any of these BSs and M nodes (e.g., M1, M2, M3, and M4) is a good candidate for placing a MEC PoP.

To better understand which location is most suitable, we need to characterize the RTT. To that end, all the network links are assumed to be fiber optic and are characterized by a propagation delay of $5 \mu s/km$ [21]. We also consider a processing delay of $50 \mu s$ on each of the M nodes [22], [23]. Therefore, the Eq. (26) becomes:

$$RTT = 2d \cdot 5 \mu s/km + 2M \cdot 50 \mu s + UL + DL$$

(28)

where $d$ stands for the distance in kilometers between the MEC PoP and the BS, and $M$ is the number of M nodes (i.e., number of hops) being traversed. For instance, $M = 0$ in case of collocating the MEC PoP with the BSs, $M = 1$ in case of collocating the MEC PoP with M1 nodes, $M = 2$ in case of collocating the MEC PoP with M2 nodes, etc. Therefore, the first two terms in the right hand side of Eq. (28) correspond to the propagation delay RTT and the packet processing delay RTT in the transport network, respectively. The last two terms (i.e., UL and DL) correspond instead to the Uplink and Downlink delay over the radio link.

3GPP defines multiple profiles for the radio interface (i.e., New Radio – NR) and each of these profiles is characterized by distinct UL and DL delay values [24]. Bound to the most stringent one-way latency of 0.5 ms for the tactile interaction URLLC traffic flow (see Table I), the BSs used for the results exposed in Sec.V are 5G gNodeBs with the suitable radio profiles that satisfy $UL + DL < RTT = 1 ms$; hence “BS” refers to a 5G gNodeB from now on. Table II reports the NR profiles used, which all adopt an uplink semi-persistent scheduling (SPS), and the maximum distances $d$ from a BS to a MEC PoP.

**Fig. 4: Reference network infrastructure as illustrated in [14] and based on [20]**
outer disk, \( D(x, \frac{a}{2}, \frac{b}{2}, a) \), where \( \|\cdot\|_V \) denotes Vincenty’s distance [26] using the WGS-84 datum [27].

For Madrid city center we consider as population circles the different districts inside the M-30 ring highway, which administratively identifies the city center and its outskirts. Then, for each of these districts we obtain the \( \rho, a, b \) values from the city center population census [28]. For what concerns the Cobo Calleja industrial area, around 100000 people \(^3\) work daily in the area, and one population circle is enough to cover the 8 km\(^2\) region. In Madrid city center and Cobo Calleja, each population circle center \( c_i \) lies over a LTE tower retrieved from OpenCellID [30]. Finally, regarding the Hoces del Cabriel rural area, rather than using the population to determine the BSs’ location, we limit our analysis to the location of 1 BS/km [14] along the A-3 highway crossing the rural area.

2) Characterization of BSs’ generation: to generate the BSs’ locations, this work uses the \textit{inhomogeneous Matérn II} point process described in Proposition III.2 together with the average number of BSs described in Sec. IV-A. In the case of Madrid city center and Cobo Calleja areas, we require an average of \( \mathbb{E}(N(R_i)) = 12 \) BSs/km\(^2\) [14] \(^4\), which is obtained by taking \( x_{1,s} = x_{2,s} = 1 \) km, \( r = 2 \cdot 1 \) km/\( \sqrt{12} \) (see Eq. (25)) and \( \lambda(x) = k \cdot G(x) \), where \( k = 16 \) for Madrid city center and \( k = 13 \) for Cobo Calleja. These values have been obtained using the average number of points expression that we derive in Proposition III.2.

The urban scenario necessitates additional considerations on the indoor hotspot traffic flow, which is not ubiquitous but rather present at few and specific location in Madrid city center. Following the same approach of [14], we consider the indoor hotspot traffic flow to be present only in office buildings \(^5\) which require 4 femtocells on each floor.

On the other hand for Hoces del Cabriel Valley, as mentioned in Section IV-A, we locate 1 BS/km along the A-3 highway to support the intelligent transport system flow. That is, the location and number of BSs follows the route of the highway rather than the population.

3) Characterization of MEC PoPs maximum distances: Algorithm 1 uses Eq. (27) to determine the maximum separation between a MEC PoP and a BS. To do so it needs to know the used distance for the propagation delay between a MEC PoP and a BS (i.e., what is \( d \) in \( l(\|x - m\|_d) \)), and the used BSs’ NR profile of Table II. Since Sec. IV-A assumes that a BS is connected to a MEC PoP with a fiber link, which are usually installed along the road lanes, Algorithm 1 uses the Manhattan distance, so we have \( l(\|x - m\|_d) = l(\|x - m\|_1) \). Regarding the NR profile, it assumes that all the generated BSs have the same radio technology to have a fixed value of \( t_e = UL + DL \) in Eq. (27). Hence we need a dedicated execution per NR profile to know how it affects the MEC PoPs’ deployment.

\(^3\)Estimation based on the number of employees working in Cobo Calleja companies [29].

\(^4\)Here we are not considering the indoor hotspot traffic flow for the urban scenario.

\(^5\)In this work we consider office buildings with more than 15 floors [31].

For sake of clarity, when applying our model to the urban and rural scenarios, it might happen that some of the generated points are not be physically suitable for hosting a gNB (e.g., they fall in the middle of a road). Nevertheless, given the propagation delay of 5 \( \mu s/km \) for fiber optics, a slight misplacement of gNodeBs is negligible since it would only vary the end-to-end delay of few microseconds, which is an order of magnitude smaller when considering an overall latency of milliseconds.

V. SIMULATIONS AND RESULTS

To obtain the MEC PoP locations we run 100 simulations using R and the spatstat package [32]. Each simulation\(^ 6\) consists of two steps:

1) Generation of 12 BSs/km\(^2\) using “\textit{inhomogeneous Matérn II}” PPs with the parameters derived from Sec. IV-B2 (in Madrid city center the indoor hotspot BSs are included as well);

2) Generation of MEC PoPs using the NR profiles of Table II, and Algorithm 1;

For the rural scenario of Hoces del Cabriel valley, we skip step 1) and manually generate 1 BS/km along the A-3 highway that crosses the region. Therefore, we only run step 2) of the simulation to obtain the MEC PoPs needed for the BSs generated across the highway.

For each scenario only one of the 100 simulations is depicted in Fig. 5, where we represent the geographical locations of the MEC PoPs as squares or romboids depending on whether they are associated to network ring M2 or M1, respectively. Urban, industrial and rural scenarios are illustrated in Fig. 5a, Fig. 5b, and Fig. 5c, respectively; with left and right figures representing how MEC PoP locations vary depending on the used radio technology. A heat map is then used to show the number of BSs \( C_{AM} \) that can be assigned to a MEC PoP at level M. The darkest area in the rhs of Fig. 5b means that any MEC PoP associated to M2 and deployed inside that area has \( C_{AM} = 80 \) BSs whose traffic can be assigned to itself. The average number of MEC PoPs for each scenario is reported in Table III. Results show that collocating the MEC PoPs with the BSs doesn’t provide enough advantages in terms

\(^6\)Code available at: https://github.com/MartinPJorge/mec-generator/tree/32513cb7fa2ec3c22567a944d234dc6dd051a36

<table>
<thead>
<tr>
<th>( R )</th>
<th>NR</th>
<th>M1 MEC PoPs(^ 1)</th>
<th>M2 MEC PoPs(^ 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban</td>
<td>FDD 120 kHz 7s</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>FDD 30 kHz 2s</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>TDD 120 kHz 7s</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Industrial</td>
<td>FDD 120 kHz 7s</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>FDD 30 kHz 2s</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TDD 120 kHz 7s</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^ 1\) For Urban and Industrial, average number of MEC PoPs across the 100 simulations.

\( a \)
(a) Urban scenario (Madrid city center).

(b) Industrial scenario (Cobo Calleja).

(c) Rural scenario (Hoces del Cabriel).

Fig. 5: MEC PoP locations for FDD 120 kHz 7s on the left. Locations for FDD 30 kHz 2s and TDD 120 kHz 7s on the right. The markers shape indicates the network ring a MEC PoP is associated to. For each coordinate, heat map $C_{A,M1}$ denotes the number of BSs that can be assigned to a MEC PoP deployed in the target location and assigned to network ring M1. Similarly, heat map $C_{A,M2}$ denotes the number of BSs that can be assigned to a MEC PoP deployed in the target location and assigned to network ring M2.
of BSs aggregation and target traffic delay requirement. This is because the network delay (see Eq. (28)) can be satisfactorily fulfilled by aggregating more BSs in fewer MEC PoPs at higher network rings (e.g., M1, M2, etc.). In fact, Algorithm 1 minimizes the number of MEC PoPs whilst fulfilling the traffic requirements. Such traffic requirements (e.g., 1 ms RTT constraint for the URLLC slice) are never satisfied when the MEC PoPs are located at the M3 and M4 network rings, yielding to empty matrices[M3] and matrices[M4]. The reason of such behavior is that packet processing delay (see Eq. (28)) increases linearly with the number of network rings to be traversed. As a result, the MEC PoPs have been always associated to M1 or M2 network rings in all our simulations, as it can be appreciated in one of the simulation realizations shown in Fig. 5.

The lower the packet processing time, the higher the maximum distance between a BS and a MEC PoP (see Eq. (28)). Thus MEC PoPs associated to M1 have more candidate BSs to be assigned than MEC PoPs associated to M2. But among all the candidate BSs it can only have 6 BSs assigned, while a MEC PoP associated to M2 can have up to 144 BSs assigned. For both the urban and industrial scenarios, the results of our 100 simulations (see Table III) show that most of the MEC PoPs are associated to M2. Since both scenarios have short distances and propagation delays because of the high density of BSs/km², the addition of M2 packet processing delay does not exceed the 1 ms RTT of URLLC. Therefore, Algorithm 1 associates the MEC PoPs to the M2 network ring, and assigns them as many BSs as possible to reduce the number of MEC PoPs. Conversely, looking to Figure 5c, more MEC PoPs are associated to M1 in the rural scenario because distances and propagation delay to BSs are high enough to exceed the 1 ms RTT when MEC PoPs are associated to M2.

Regarding the different NR profiles (see Table II), low UL and DL delays of FDD 120 kHz 7s permit to increase the maximum distance between a BS and a MEC PoP. Therefore, a MEC PoP can serve a larger number of BSs, as shown in the darker heat maps at the lhs of Fig. 5. As a result, using FDD 120 kHz 7s as NR profile necessitates the deployment of fewer MEC PoPs compared to all the other NR profiles (see Table III). Indeed, any MEC PoP location in the urban and industrial scenario can serve any FDD 120 kHz 7s BSs in the region as shown in Fig. 5a and Fig. 5b. Instead, FDD 30 kHz 2s and TDD 120 kHz 7s impose a higher UL and DL delay, thus requiring a shorter distance between a BS and a MEC PoP. This results in a larger number of MEC PoPs sparsely located in the region (see rhs of Fig. 5).

Fig. 6 illustrates the experimental Cumulative Density Function (eCDF) for the number of BSs assigned to a MEC PoP. Results show that the FDD 120 kHz 7s NR technology increases the number of BSs associated to a MEC PoP (less than the 33% of them have less than 100 BSs assigned in the industrial and urban scenario), while the other NR technologies lead to a higher percentage of MEC PoPs with fewer BSs assigned. For example, 68% of the MEC PoPs of the industrial scenario have less than 70 BSs assigned when TDD 120 kHz 7s or FDD 30 kHz 2s are used, which is less than half the BSs that can be assigned to a MEC PoP associated to M2.

Summarizing, there is a trade-off between the performance of the NR profile and the number of MEC PoPs. Higher performance radio profiles, which can be more expensive, allow to associate a larger number of BSs to a MEC PoP, resulting in fewer MEC PoPs. On the contrary, a less performant radio profile, which is cheaper, requires a large number of MEC PoPs for satisfying URLLC traffic. This trade-off should be taken into consideration by the network operators to optimize costs when building their network.

VI. CONCLUSIONS AND FUTURE WORK

In this article we have presented a mathematical model to determine the deployment of Base Stations (BSs) and MEC Points of Presence (PoPs), using novel point processes that account for both people population and minimum distances between BSs. The model is applied in real urban, industrial and rural scenarios, where we generate 5G gNodes B and MEC deployments that satisfy the strictest 5G latency constraint of Ultra-Reliable and Low Latency Communications (URLLC) slices, and thus can support future Augmented Reality/Virtual Reality (AR/VR) and low latency streaming services. We have also analyzed the use of 3 future New Radio (NR) profiles of 5G and simulations show that FDD 120 kHz 7s is the technology that minimizes the number of MEC PoPs needed to support the future traffic demands. Future directions of this work include the analysis of the resource requirements of the MEC PoPs, the usage of clustering techniques for the assignment of BSs to MEC PoPs, and the formulation of an optimization problem to minimize MEC PoP deployment costs.

REFERENCES


