Modelling User Radio Access in Dense Heterogeneous Networks

Modelling User Radio Access in Dense Heterogeneous Networks

Marco Gribaudo[a], Daniele Manini[*,b], Carla Fabiana Chiasserini[c]

[a] *Politecnico di Milano, Milano, Italy*
[b] *Università di Torino, Torino, Italy*
[c] *Politecnico di Torino, Italy, CNIT, Italy, and CNR-IEIIT, Italy*

Abstract

One of the distinctive features of today's mobile networks is the densification of the access nodes and their heterogeneity, which lead to complex, multi-tier, multi-radio access systems. Unlike previous work, which has focussed on optimal techniques for user assignment and technology selection schemes, in this paper we present a flexible analytical model for the performance evaluation and the efficient design of the above complex systems. Leveraging a Markovian agent formalism, the model captures several essential elements, including the spatial and temporal dynamics of the user traffic demand and the availability of radio resources. Importantly, the model exhibits low complexity and an excellent match with simulation results; furthermore, it is general enough to accommodate various network architecture and radio technologies. Through an innovative mean-field solution, we derive a number of relevant performance metrics and show the ability of our framework to represent the system behavior in large-scale, real-world scenarios, with time-varying user traffic.

Key words: Performance evaluation and modelling, heterogeneous wireless networks, resource allocation.

1. Introduction

One of the essential factors driving the design of mobile networks is the need to satisfy the increasing demand for large and fast data transfers. Examples of Enhanced Mobile Broadband (eMBB) applications include on-line gaming, high-definition video, and emerging applications like augmented and virtual reality. Recent data confirm these trends: according to the Ericsson's Mobility Report 2019 [1], in 2024 5G networks will carry 35% of the mobile data traffic, and the latter is predicted to reach 131 exabytes per month with a major contribution of the aforementioned applications.

Since most of the traditional bands (namely, frequencies in the 300 MHz-3 GHz range) are already allocated to services and capacity has almost reached Shannon's

[*] Please address correspondence to Author3

Email addresses: gribaudo@elet.polimi.ita (Marco Gribaudo), manini@di.unito.it (Daniele Manini), chiasserini@polito.it (Carla Fabiana Chiasserini)

limit [2], a relevant solution to the throughput demand by data-hungry applications is to develop heterogeneous, dense communication networks [3]. Such networks are characterized by the presence of large cells and many small cells, with the latter ones being classified as micro-cells, pico-cells, or femtocells [4]. While femtocells are mostly used indoors, micro and macro cells are already essential components of outdoor cellular networks. As a consequence, their densification is pivotal to the provisioning of high data rates, to the support of a large number of simultaneous connections, and, thanks to the shorter distance between users and network Point of Access (PoA), to a reduced power consumption.

Such multi-tier network architecture will necessarily leverage multiple radio access technologies (RAT), like LTE, WiFi, and mmwave, just to name some of the most relevant user-infrastructure communication technologies. Several studies have therefore studied, e.g., LTE access systems with two tiers as well as different RATs in 5G RAN systems [5], including the possibility to support the implementation of virtual RATs [6]. In this scenario, users are, and will ever be, equipped with more and more radio interfaces, thus enabling seamless connections with different PoAs and high-throughput performance anytime and everywhere. None the less, several challenges still need to be addressed, among which, the management of radio resources – a key factor to the effective design of heterogeneous networks.

In this paper, we address the above challenge and develop a framework for the analysis of strategies to effectively handle the connectivity of mobile users. In particular, we consider that the network service area is covered by a number of PoAs, each of them hosting multiple radio access interfaces. A user can connect to any of such interfaces, provided that it is under coverage and enough radio resources are available. To efficiently allocate the radio resources, we propose an analytical model, which leverages mean-field analysis [7, 8]. Unlike other techniques, such as queueing networks, stochastic Petri nets, or process algebras, mean-field analysis permits to account for the spatial distribution of the communication nodes (PoA and users) in the system. This is clearly of fundamental importance, as a user can access a PoA only if its position is within the coverage area of that PoA. Importantly, the model exhibits a low complexity, hence it is able to deal with very large network scenarios. The model is then solved by resorting to a novel method based on the Markovian Agent formalism [9], and exploiting the results in [10].

For sake of concreteness, we focus on a specific, fully-distributed, user-centric strategy for RAT selection and study the system performance in a real-world scenario. It is worth remarking that our analytical framework can be extended to investigate other RAT selection strategies, as well as different heterogeneous, multi-layer scenarios. The main contribution of this work resides in the development of an effective and efficient application model. Furthermore, the paper provides relevant, theoretical insights through a set of equations that, by a mean field-based model, capture the dynamics of a RAT communication network.

The rest of the paper is organised as follows. In Section 2, we discuss some related work on the modelling and analysis of new generation networks. Section 3 first introduces the system scenario and the assumptions we make on both the communication network and the user traffic demand; then it provides an example of user access policy in a multi-technology network. Section 4 describes the model we developed to

represent the heterogeneous network system and its model parameters, while Section 5 presents the solution techniques we adopted to solve our model. Section 6 introduces the network topology and the instance of the general model we used for validating our approach, and it shows a comparison between analytical and ns3 simulation results. Section 7 applies the proposed model and solution technique to a real-world scenario, leveraging both user mobility and traffic demand experimental traces. Finally, Section 8 draws some conclusions.

2. Related Work

It is foreseen that 5G networks allow concurrent association of users with different network types of infrastructure, exploiting the multihoming feature of mobile devices and seamless connection switch from one network PoA to another within the same Radio Access Network (RAN), or between different RANs [11, 12]. Concurrent user association and traffic switching between different RANs need to be performed taking into account the characteristics of each RAN as well as the application requirements [13, 14]. To this end, several studies have investigated handover mechanisms in heterogeneous networks [15, 16]. Additionally, standards like IEEE 802.21 leverage interoperability between heterogeneous networks for handover optimization [17], without, however, specifying any network selection mechanisms.

The problem of determining the optimal user association in the presence of different radio access technologies has also received a great deal of attention. From the theoretical viewpoint, two main approaches are usually adopted: game theory and optimization exploiting the utility concept. The former aims at formulating a game where convergence to a Nash equilibrium and the Pareto-efficiency of this equilibrium can be provided [18, 19, 20]. The latter, instead, foresees that the network corresponding to the highest utility value is selected: the utility can be a function of monetary cost, power consumption, network conditions, or user preferences [21, 22].

In the specific scenario where WiFi and cellular technologies coexist, the work in [23] presents a cross-technology communication mechanism for efficient resource allocation and interference management, while optimal coexistence mechanisms and radio selection schemes have been proposed, e.g., in [24] and [25], respectively. In [26], a urban deployment scenario is investigated, where WiFi small cells are overlaid on top of the 3GPP LTE network. The authors propose user-centric network selection algorithms to minimize feedback overhead while taking into account user preferences. A comprehensive review of the state-of-the-art mathematical models that have been applied to the network selection problem, including utility-based approach, game theory, combinatorial optimization, Markov decision processes, and fuzzy logic, can be found in [27].

While several experimental works exist, which show the performance of multiple radio-access technologies and front-haul systems [28, 29, 30, 31, 23], fewer studies have focussed on to the development of theoretical models of a heterogeneous, multi-radio access mobile network, allowing for the performance evaluation of large-scale systems. Among these, the work in [32] addresses a multitier, heterogeneous scenario and presents a matching theory-based algorithm and a scheduler that maximize the

number of users served in a reliable manner. The study in [33] introduces an AI-based framework to facilitate an effective resource management. An approach based on stochastic geometry is instead used in [34] to model dual connectivity and user-centric virtual cells, as well as to derive expressions for, e.g., the coverage probability of a typical user in downlink cellular.

In [35], a Markovian technique is used to study wireless sensor networks, showing the applicability of this approach in a quite complex, real scenario. The modelling of virtualized network functions in 5G networks is addressed in [36]. Interestingly, mean-field solutions have been recently improved in [37] to obtain accurate results, even in systems with a small population. Markovian Agents and mean-field analysis have been used in a similar way to the one adopted in this work to study wireless sensor networks, in [38] and [39], respectively. Finally, we mention that a similar methodology, although simpler and of less general validity, was introduced in our conference papers [40, 41].

Novelty. In this work, we propose a flexible, analytical model capturing the users' as well as PoAs' geographical locations, the availability of different radio resources, the spatial and temporal dynamics of the user traffic demand, and the features of a multi-tier network architecture with overlapping PoA coverages. The model, based on Markovian agents, is general and can accommodate different radio access technologies, various types and densities of small cells, and diverse user traffic patterns. Importantly, we use an efficient and neat solution technique, which is first applied to the field of heterogeneous networks, and we derive several relevant system performance metrics. Since the proposed approach takes advantages of mean-field analysis, it is possible to analyse large-scale systems. Indeed, we show the effectiveness of our methodology by applying it to a large-scale, real-world scenario. Additionally, the technique is validated by simulation using ns3.

3. System Scenario

In this section, we first introduce the system scenario and the assumptions we make on both the communication network under study and the user traffic demand (Section 3.1). We then provide an example of user access policy in a multi-technology network, which we will adopt for concreteness while presenting our analytical model (Section 3.2).

3.1. Network scenario

Let $\mathcal{V}$ be a urban area under the coverage of a multi-tier 5G network characterised by multi-technology radio access. Let us then focus on two sets of cells, namely, macrocells and microcells, although the model can be easily extended to the case where more cell tiers are present, from femtocells to macrocells. Macrocells are served by PoAs such as cellular base stations operating at relatively low frequencies in the licensed spectrum (e.g., today's LTE frequency bands), while microcells are served by PoAs operating at higher frequencies and use WiFi or mmWave technologies. Thus, independently of their technology and location, cells are assigned disjoint sets of downlink radio resources, i.e., there is no inter-cell interference between any two PoAs.

We denote the number of macrocells by N_M, the generic macrocell by M_i, $i = 0, \ldots, N_M - 1$, and the coverage area of the generic macrocell M_i by $\mathcal{A}_i$. We then

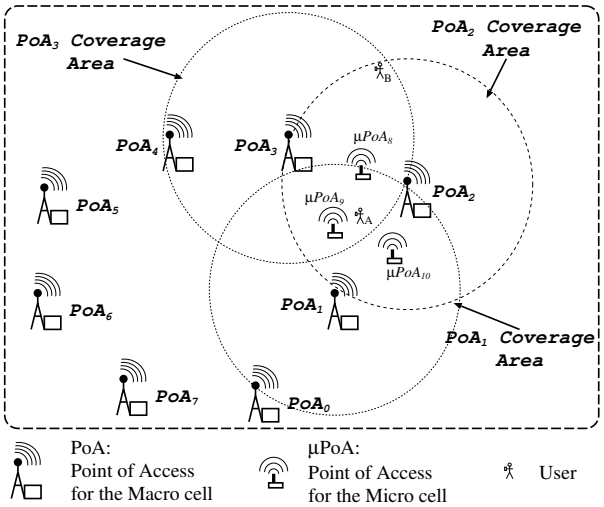


Figure 1: An example of multi-tier network topology: PoAs covering macro-cells and μPoAs covering microcells are sequentially numbered (the former ones range from 1 to 7, and the latter from 8 to 10).

assume that area $\mathcal{A}_i$ ($i = 0, \ldots, N_M - 1$) includes N_{mi} microcells, each denoted by m_p and with coverage area a_p ($p = 0, \ldots, N_{mi} - 1$). Coverage areas of neighbouring macrocells and microcells may overlap. Let $\mathcal{H}_i$ be the set of macrocells that are neighbours of M_i, and by h_p the set of microcells that are neighbours of m_p. Users located in area $\mathcal{A}_i \cap \mathcal{A}_j$ can access either M_i or M_j; similarly, users in $a_p \cap a_q$ can access either m_p or m_q. A graphical representation of the network scenario is depicted in Fig. 1, where $N_M = 8$ and, e.g., $\mathcal{H}_2 = \{3, 1\}$ and $h_9 = \{8, 10\}$.

Since data traffic in wireless networks is still highly asymmetric [42], we focus on downlink data transfers (i.e., from the PoA to the users), such as content downloading or video streaming, and denote by N_W the number of radio resources available in each cell. Importantly, according to the heterogeneous 5G radio interfaces, radio resources can be defined in terms of frequency, time, antennas, or transmit power.

As mentioned, users that are under the coverage of more than one PoA can connect to any of them, however they will likely experience different levels of link quality depending on the PoA they receive service from. In spite of the above, it is fair to assume that, for all technologies, the link quality decreases as the distance between a user and a PoA increases, and the better the link quality, the higher the transmission rate that the link end points can use.

Finally, since we are interested in studying traffic offload/onload dynamics between different network tiers, it is fair to focus on pedestrian users, which can be served by microcells or macrocells depending on their location. It follows that, since a data transfer from the Internet is expected to last just few seconds, users do not move significantly during a data traffic transfer.

3.2. Network selection strategy

We recall that macrocells and microcells use different portions of the spectrum, and, possibly, even licensed and unlicensed frequencies, respectively. Thus, accessing

a macrocell PoA or a microcell PoA implies a different monetary cost for the user. For example, the access to a WiFi microcell may have little or no cost, unlike the access to a mmWave cellular microcell or to a macrocell, which however can offer throughput guarantees. It is therefore sensible to envision an example selection strategy, according to which users access the different tiers of the network.

For sake of concreteness, we specify one of such policies, which aims at ensuring that, for each newly initiated traffic transfer, a user selects the technology tier (there may be two or more the user can connect to) providing a sufficiently high throughput while implying little monetary access cost. We remark that the model and the solution we present in this paper can be extended to any other selection strategy that may be implemented in a heterogeneous network, with slight modifications to the appropriate functions introduced in Section 4.

The example strategy we envision is fully distributed, i.e., users independently decide which technology to use in order to download a content from the Internet and they cannot relay on detailed, a priori information on the throughput they will experience when connected to an access interface. More specifically, we consider that a user lists the available PoAs, i.e., those PoAs from which it detects a sufficiently high received signal strength (RSS), from the strongest to the weakest. Then (i) if any microcell appears in the list, it connects to the one from which it receives the strongest signal (hereinafter referred to as *local microcell*), (ii) otherwise it connects to a macrocell (referred to as *local macrocell*). In the case where a microcell is selected (case (i)), if the experienced throughput is below a given threshold μ for a sufficiently long time (hysteresis time[1]), the user switches either to an available microcell (named *neighbouring microcell*), or to a local macrocell PoA, from which it receives a better service. Depending on the experienced throughput, the user migrates again if the performance is below the μ threshold for a hysteresis time and the traffic transfer has not yet been completed. Note that when a user is served by a local macrocell, it can only migrate to a neighbour macrocell.

When instead a user is under coverage of macrocell PoAs only (case (ii)), it connects to its local macrocell, and it switches to the neighbouring macrocell only if it experiences an unsatisfactory throughput. Furthermore, a user receiving data from a neighbouring macrocell periodically tries to migrate to its local macrocell.

We remark that such a selection policy can be easily extended to the case where more network tiers are available, and, as also specified later, the analytical framework we develop can easily accommodate other policies as well.

4. A Markovian Agent Model

We now describe the model we developed to represent the heterogeneous network system and its model parameters, highlighting how our model can capture both the multi-tier communication and the user traffic demand characterizing a given geographical area. More in detail, after introducing the notations used in Section 4.1, we model the user possible states through a continuous time Markov chain in Section 4.2. Therein

[1] A hysteresis time is foreseen in order to avoid the well-known ping-pong effect.

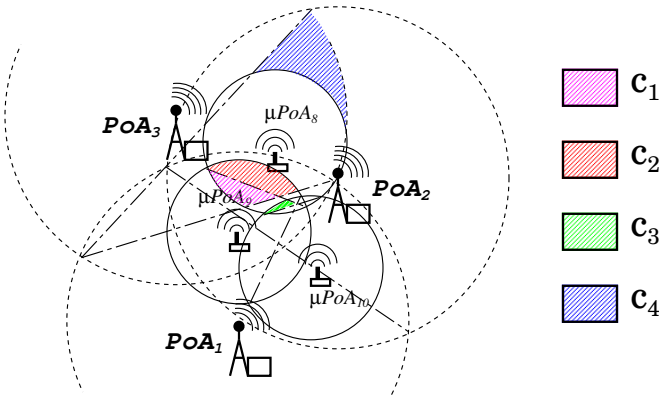


Figure 2: Model classes and coverage areas.

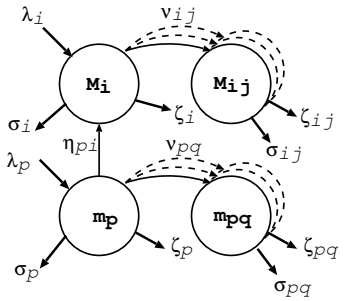


Figure 3: Mean-field model of a PoA: agents' states and state transitions.

we also focus on high user-density scenarios, and adopt a mean-field approximation and a counting process to derive relevant performance metrics. Finally, in Section 4.3, we characterize the functional rates determining the system evolution.

4.1. Notations

We indicate vectors and matrices in bold, with lowercase and capital letters, respectively; matrix and vector elements are instead denoted by lowercase letters. Sets are denoted with calligraphic uppercase letters. Uppercase roman letters indicate constants, while greek letters denote functional rates. Finally, the name of the states, in which the system can be, is denoted with typewriter letters. All system and model notations are summarized in Table 1.

4.2. System model

We develop a Markovian Agent Model (MAM) where each user wishing to download data is represented by a Markovian Agent (MA). More specifically, agents are spread over the geographical region $\mathcal{V}$ and each of them is described by a Continuous Time Markov Chain (CTMC) whose infinitesimal generator depends on the MA structure and the user's geographical position $\mathbf{v_i} \in \mathcal{V}$.

Table 1: Notations

System Notation	Description
N_M	number of macrocells
M_i	generic macrocell i
$\mathcal{A}_i$	coverage area of macrocell M_i
$\mathcal{H}_i$	set of macrocells that are neighbours of M_i
N_{mi}	number of microcells in $\mathcal{A}_i$
m_p	generic microcell p in $\mathcal{A}_i$
a_p	coverage area of microcell m_{ip}
h_p	set of microcells that are neighbours of m_p
N_W	number of available radio resources
S	technology throughput speed
F	throughput reduction constant
μ	minimum throughput threshold
N_Q	maximum number of connections

Model Notations	Description
$\mathcal{V}$	geographical region
$\mathbf{v_i}$	geographical position in $\mathcal{V}$
N_c	number of agents' classes
c	generic class
st_c	number of states of class c
m_p	state *local microcell*
M_i	state *local macrocell*
m_{pq}	state *neighbour microcell*
M_{ij}	state *neighbour macrocell*
L_i	state *loss* for PoA
$n_j^{(c)}(t, \mathbf{v})$	average number of agents of a class c in state j at time t at location $\mathbf{v}$
$\rho^{(c)}(t, \mathbf{v})$	average number of agents of class c in location $\mathbf{v}$ at time t
$p_j^{(c)}(t, \mathbf{v})$	probability of state j
$[N_{\mathcal{V}}]$	average number of agents of all classes and locations
$\mathbf{b}^{(c)}(t, \mathbf{v}, [N_{\mathcal{V}}])$	birth term
$\mathbf{K}^{(c)}(t, \mathbf{v}, [N_{\mathcal{V}}])$	transition kernels
$\mathbf{Q}^{(c)}(t, \mathbf{v})$	local transitions matrix
$\mathbf{I}^{(c)}(t, \mathbf{v}, [N_{\mathcal{V}}])$	influence matrix
$\mathbf{D}^{(c)}(t, \mathbf{v})$	death term
$d_{ii}^{(c)}(t, \mathbf{v})$	death term element (death rate)
$\lambda_{m_p}(t)$	arrival requests rate in a_p at time t
$\lambda_{M_i}(t)$	arrival requests rate in $\mathcal{A}_i$ at time t
σ_x	service rate in state x
p	probability of horizontal handover
$(1-p)$	probability of vertical handover
μ_x	minimum throughput threshold in state x
ν_x	horizontal handover rate in state x
η_x	vertical handover rate in state x
r_x	switching rate in state x
$u(z)$	indicator function
$c_{sigmoid}$	constant of the Sigmoid function
ζ_x	loss rate for local users in state x
ζ_x^*	loss rate for remote users in state x
$N_T(\mathcal{A}_i)$	number of technologies in area $\mathcal{A}_i$

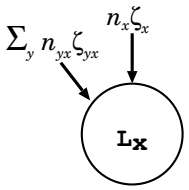


Figure 4: Mean-field model of traffic losses at a PoA.

Agents can be classified based on the connectivity opportunities available to the users they represent as well as on the users' traffic demand. As an example, users under the coverage of a macrocell only, have different connectivity opportunities from those under the coverage of both macro and microcells. Denoting with N_c the number of agents' classes, an MA belonging to class c ($1 \leq c \leq N_c$) is characterised by st_c different states, which in turn model the coverage conditions of the corresponding user. Fig. 2 exemplifies four different classes of agents, characterised by different user positions. Users[2] in class c_1 start connecting with μPoA_9 and can switch to μPoA_8, while users in c_2 connect initially with μPoA_8 and can then switch to μPoA_9. Both classes can be served by PoA_1, PoA_2, and PoA_3, with PoA_2 being the closest one. Users in c_3 start connecting with μPoA_{10}, while the ones in c_4 cannot be served by any microcell.

To summarise, a class is an ordered set of PoAs, at different locations and of different type, to which a user can connect according to the given network selection policy. It follows that all the users located in the same restricted area, e.g., the coloured shapes in Fig. 2, belong to the same class, and the users of a specific PoA are spread over several classes. The shapes of these restricted areas roughly correspond to the intersection of the PoA coverage areas. Note that any user only belongs to one class according to her/his position.

The agent's states are as follows: m_p (*local microcell*), M_i (*local macrocell*), $m_{pq} \in h_p$ (*neighbour microcell*), $M_{ij} \in \mathcal{H}_i$ (*neighbour macrocell*), which account for any local or neighbouring microcell and macrocell the user can connect to. Fig. 3 reports a meta-model of a generic class that can be instantiated according to the number of neighbours and available technologies for a user under the coverage of a macrocell M_i and a microcell m_p. Furthermore, to effectively capture the mobile network scenario, we assume that the traffic demand depends on both the agent position $\mathbf{v_i} \in \mathcal{V}$ and the current time. To account for the fact that the user's traffic request may not be served by any PoA, e.g., due to lack of resources or poor channel conditions, we introduce the *loss* agent characterised by a single state L_x ($x = p, i$), as shown in Fig. 4.

Next, focusing on the most challenging case where the communication network has to serve a large number of users, we exploit a mean field approximation and use a *counting process* [43, 8] to determine the distribution of agents across their different states. To this end, we proceed as follows.

Let $n_j^{(c)}(t, \mathbf{v})$ be the average number of agents of a class c in state $1 \leq j \leq st_c$ at

[2]For simplicity, we refer to agents representing users also as "users".

time t and location $\mathbf{v} \in \mathcal{V}$, and let $\boldsymbol{n}^{(c)}(t,\mathbf{v}) = |n_j^{(c)}(t,\mathbf{v})|$. Then we define $\rho^{(c)}(t,\mathbf{v})$ as the average number of agents of class c in location $\mathbf{v}$ at time t (i.e., not considering the connectivity opportunities), and $p_j^{(c)}(t,\mathbf{v})$ as the corresponding probability of state j. We can then write:

$$n_j^{(c)}(t,\mathbf{v}) = p_j^{(c)}(t,\mathbf{v}) \cdot \rho^{(c)}(t,\mathbf{v}). \tag{1}$$

with:

$$\begin{aligned} \rho^{(c)}(t,\mathbf{v}) &= \sum_j n_j^{(c)}(t,\mathbf{v}) \\ p_j^{(c)}(t,\mathbf{v}) &= \frac{n_j^{(c)}(t,\mathbf{v})}{\rho^{(c)}(t,\mathbf{v})}. \end{aligned}$$

To determine $\boldsymbol{n}^{(c)}(t,\mathbf{v})$, let us define $[N_\mathcal{V}] = \{(c,\mathbf{v},\boldsymbol{n}^{(c)}(t,\mathbf{v})) : 1 \le c \le N_c, \mathbf{v} \in \mathcal{V}\}$ as the average number of agents in the ensemble of all classes and all locations at time t. The temporal evolution of $\boldsymbol{n}^{(c)}(t,\mathbf{v})$ is described by the following differential equation:

$$\frac{d\boldsymbol{n}^{(c)}(t,\mathbf{v})}{dt} = \mathbf{b}^{(c)}(t,\mathbf{v},[N_\mathcal{V}]) + \boldsymbol{n}^{(c)}(t,\mathbf{v}) \cdot \mathbf{K}_c(t,\mathbf{v},[N_\mathcal{V}]). \tag{2}$$

Term $\mathbf{b}^{(c)}(t,\mathbf{v},[N_\mathcal{V}])$ is the so-called *birth* term and expresses the rate (measured in agent density per time unit) at which class c agents are created in location $\mathbf{v}$ at time t. The term $\mathbf{K}^{(c)}(t,\mathbf{v},[N_\mathcal{V}])$ is the *transition* kernel, which determines the evolution of the system: it accounts for both the state transitions of the agents and for the effects that the number of agents in one location may decrease. In general, it depends on the class c, the position $\mathbf{v}$, the time t, and the ensemble average count $[N_\mathcal{V}]$. More specifically, the transition kernel includes three terms:

$$\begin{aligned} \mathbf{K}^{(c)}(t,\mathbf{v},[N_\mathcal{V}]) &= \mathbf{Q}^{(c)}(t,\mathbf{v}) + \mathbf{I}^{(c)}(t,\mathbf{v},[N_\mathcal{V}]) + \\ &\quad -\mathbf{D}^{(c)}(t,\mathbf{v},[N_\mathcal{V}]) \end{aligned} \tag{3}$$

where:

- for every agent class c and position v, matrix $\mathbf{Q}^{(c)}(t,\mathbf{v}) = |q_{ij}^{(c)}(t,\mathbf{v})|$ defines the rate of local transitions at time t, as in traditional CTMC. In practice, it represents autonomous changes of state, i.e., non induced by the state of other agents;
- the influence matrix $\mathbf{I}^{(c)}(t,\mathbf{v},[N_\mathcal{V}])$ accounts for the rate of *induced transitions*, due to the influence of other agents. The entries of matrix $\mathbf{I}^{(c)}(t,\mathbf{v},[N_\mathcal{V}])$ depend on the average number of agents in neighbouring locations, and are defined such that matrix $\mathbf{Q}^{(c)}(t,\mathbf{v}) + \mathbf{I}^{(c)}(t,\mathbf{v},[N_\mathcal{V}])$ is still an infinitesimal generator matrix;
- $\mathbf{D}^{(c)}(t,\mathbf{v})$ is the so-called death term; it is a diagonal matrix whose generic element $d_{ii}^{(c)}(t,\mathbf{v})$ represents the rate at which a class c agent at location $\mathbf{v}$ and in

state i moves out from the system at time t. In our scenario, it can correspond to either a service or a loss.

To solve the above model, the initial state of the system must be provided, namely, the initial density of class c agents in location $\mathbf{v}$, $\rho^{(c)}(0, \mathbf{v})$, and the corresponding initial state probability, $p_j^{(c)}(0, \mathbf{v})$. We can then compute the initial condition of (2), as:

$$n_j^{(c)}(0, \mathbf{v}) = p_j^{(c)}(0, \mathbf{v}) \cdot \rho^{(c)}(0, \mathbf{v}). \tag{4}$$

Furthermore, matrices $\boldsymbol{\nu}(t, \mathbf{v}, [N_{\mathcal{V}}])$ and $\mathbf{K}^{(c)}(t, \mathbf{v}, [N_{\mathcal{V}}])$ have to be particularized so as to represent the specific network system under study.

Importantly, once we compute the time evolution of $n_j^{(c)}(t, \mathbf{v})$, we obtain the average number of users over time, for each possible connectivity opportunity and spatial location, i.e., the number of users connected to a generic PoA. The average number of users accessing a PoA is indeed of crucial importance, since, along with the knowledge of the radio resources available at the PoA and the adopted traffic scheduling policy, it allows us to compute the resource share per user. Then, given the experienced channel quality, the data rate enjoyed by the user can be readily obtained. Finally, using the user's data rate and the user's location, the next state of the corresponding agent can be derived according to the network selection strategy discussed in Section 3.2.

Fig. 3 depicts the overall model and highlights the dynamic of a single PoA: incoming arrows correspond to the arrival of user's service requests that should be accommodated, while outgoing arrows represent users leaving the PoA, i.e., users that have completed their download or connecting to another PoA. In our model, matrices and vectors illustrated above are derived by the definition of the functional rates (see Section 4.3) that describe the dynamic behaviour of the different components. The example provided in Section 5 presents how the transition rates are instantiated to obtain Equations 2 and 3.

4.3. Functional rates

The user connectivity dynamics are represented by the transition rates, which are functions of the number of users in each state representing the local and the neighbouring PoAs. We define below the functional rates in our model, accounting for an arbitrary number of neighbouring PoAs for each given PoA, as well as for time-dependent resource allocation policies. To simplify the notation, in the following we denote with $n_x = \sum_c n_x^{(c)}(t, \mathbf{v})$ the total number (over all classes) of agents in state x. Also, we no longer denote the time dependency and drop the t variable in all the equations.

Arrivals. We model the service requests per second generated by users located in coverage area a_p with the functional rate $\lambda_{m_p}(t)$ $(p = 0, \ldots, N_{mi} - 1)$, and those by users located in region $\mathcal{A}_i$ with functional rate $\lambda_{M_i}(t)$ $(i = 0, \ldots, N_M - 1)$. In this way, the generation of download requests can exhibit both a spatial and a time dependency: we will exploit this feature to generate bursts of requests at specific locations, as well as to study the corresponding network performance.

Services. The service rate of a state x, with $x = m_p, M_i$, can be expressed as

$$\sigma_x = \frac{S_x \times N_{wx}}{n_x + F_x \sum_{y \in \mathcal{X}_{\S}} n_{yx}} \tag{5}$$

where S_x is the throughput (connection speed) when the PoA is using its radio interface to serve only one user, and N_{wx} is the number of resources available. The denominator accounts for the total number of users that are sharing the radio resources, i.e., the local users n_x and the users n_{yx} accessing the considered cell from remote locations, i.e., $\mathcal{X}_x = \mathcal{H}_x, h_x$ (hereinafter also referred to as remote users).

F_x is a constant equal to 1 when $x = M_i$, and equal or greater than 1 when $x = m_p$. Indeed, when the cell is a microcell, this factor takes into account the throughput reduction that remote (hence, far away) users may experience while accessing the microcell. As an example, if the microcell technology is WiFi with IEEE 802.11a/g interface, then F_x accounts for the anomaly effect [44] typical of such technologies. Similarly, we define $\sigma_{xy} = \sigma_y$ with $xy = pq, ij$, since the corresponding agents are served by PoA_y.

Horizontal and vertical handover rates. Rates ν and η represent the horizontal and vertical handover rates, respectively, which depend on the network selection strategy. Specifically, recalling the policy in Section 3.2, if the local throughput drops below a minimum threshold μ_x, a user will switch with probability p_y to the interface of a neighbouring cell y (horizontal handover), or with probability $1 - p$ to a local macrocell (vertical handover), where $p = \sum_{y \in \mathcal{X}_x} p_y$.

The handover rates then depend on the system state, the technology latency, and the implemented switching policy. In particular, the horizontal handover rate toward a state y, with $y = m_{pq}, M_{ij}$, is given by:

$$\nu_y = p_y r_x u(\mu_x - \sigma_x) \tag{6}$$

where r_x is a system parameter accounting for the switching rate from one PoA to a neighbouring PoA, σ_x is the throughput associated with the current state (i.e., the used technology) as defined in (5), and $u(z)$ is an indicator function that returns 1 if the argument is positive and 0 otherwise. To improve the numerical stability of the solution we used the Sigmoid function below:

$$u(z) = \frac{c_{sigmoid}}{1 + e^{(-z)}} \tag{7}$$

where $c_{sigmoid}$ is a constant that controls the slope of the function.

Likewise, we define the vertical handover rate η_x, with $x = m_p$ (for simplicity, in this model we only consider the vertical handover from microcells to macrocells):

$$\eta_x = (1 - p) r_x u(\mu_x - \sigma_x)\,. \tag{8}$$

Note that the model is very flexible with respect to the network selection policy (hence, handover strategy), since the extension to other policies can be implemented

by simply changing the definition of ν and η.

Loss of requests. Losses are modelled by the ζ_x and ζ_{xy} out-coming arcs in the PoA model, and by the loss agent depicted in Fig. 4. In particular, they are defined to balance the exceeding traffic that cannot be served, i.e.,

$$\zeta_x = \sigma_x u \left(n_x + \sum_{y \in \mathcal{X}_x} n_y > N_{Qx} \right) \tag{9}$$

$$\zeta_{xy} = \sigma_y u \left(n_y + \sum_{z \in \mathcal{X}_y} n_z > N_{Qy} \right) . \tag{10}$$

In other words, as long as the total number of connections in a PoA is less than a given threshold N_{Qx}, no loss occurs, i.e., the above rates are all equal to 0. Otherwise, the number of traffic transfers exceeding such a threshold are lost and accounted in the corresponding loss states.

Note that, given a loss agent, L_x, the incoming arcs sum up all the losses for local users and remote users accessing PoA_x and located in different neighbouring cells. In particular, their rates correspond, respectively, to $n_x \zeta_x$ and $\sum_y n_{yx} \zeta_{yx}$ since every user connected to PoA_x experiences the same loss rate.

5. Model Solution

To solve the model we developed, we adopt the solution techniques outlined in [10] and proceed as follows.

Let us focus on class of users i modeling coverage area $\mathcal{A}_i$ ($i = 1, \ldots, N_c$). As the first step, the agent exemplified in Fig. 3 is used to determine the states corresponding to coverage area $\mathcal{A}_i$. In particular, if a user in class i can access $j = 1, \ldots, N_T(\mathcal{A}_i)$ technologies, each one characterised by $n_j(\mathcal{A}_i)$ neighbouring cells, the agent will be described by $\sum_{j=1}^{N_T(\mathcal{A}_i)} (n_j(\mathcal{A}_i) + 1)$ states, where "+1" represents the local cell for each technology. Let variable $n_x^{(i)}$ count the number of agents in the considered area in state x, and let row vector $\boldsymbol{n}^{(i)} = |\boldsymbol{n}_x^{(i)}|$ collect such variable related to class i.

The *transition matrix* $\mathbf{Q}^{(i)} = |q_{xy}^{(i)}|$ (see (3)), containing the transition rates $q_{xy}^{(i)}$ from state x to state y, can be obtained by using the rates defined in Section 4.3. The *death matrix*, $\mathbf{D}^{(i)} = diag(d_{xx}^{(i)})$, is a diagonal matrix whose elements $d_{xx}^{(i)}$ represent agents that leave the system in state x (in our case, they correspond to successful request transmissions). Finally, the *birth vector* $\mathbf{b}^{(i)} = |b_x^{(i)}|$ (see (2)) represents the arrival of new agents in state x. As for the transition matrix $\mathbf{Q}^{(i)}$, also the elements of $\mathbf{D}^{(i)}$ and $\mathbf{b}^{(i)}$ can be computed using the rates presented in Section 4.3. (As done in Section 4.3, we omit the explicit dependency from time and location to improve readability.)

Then the system evolution is obtained by solving for each class i (2), which, in this case, simplifies to:

$$\frac{d\boldsymbol{n}^{(i)}}{dt} = \boldsymbol{n}^{(i)} \left(\mathbf{Q}^{(i)} - \mathbf{D}^{(i)} \right) + \mathbf{b}^{(i)}. \tag{11}$$

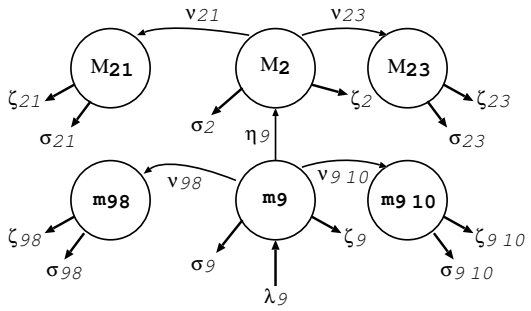


Figure 5: Example of an agent model (Agent A).

Note that, even if (11) is solved separately for each class, terms $\mathbf{Q}^{(i)}$, $\mathbf{D}^{(i)}$, and $\mathbf{b}^{(i)}$ depend on the complete state of the system, thus allowing the induction mechanism to capture the interaction between different classes. As an example, let us focus on the user A depicted in Fig. 1: A is directly connected to PoA_2 and μPoA_9, and it can switch to PoA_1, PoA_3, μPoA_8, or μPoA_{10}. Let class A denote the class modelling this user, and let m_9, m_{98}, $m_{9\ 10}$, M_2, M_{22}, and M_{23} be the order of the states in which the user can be. The resulting model is depicted in Fig. 5, while the corresponding matrices $\mathbf{Q}^{(A)}$ and $\mathbf{D}^{(A)}$, and vector $\mathbf{b}^{(A)}$ are reported in Fig. 6. Similarly, if we consider agent B in Fig. 1, which is under the coverage of PoA_2 and PoA_3, its agent is characterised by states M_2 and M_3, while its class and matrices are shown in Fig. 7.

$$\mathbf{Q}^{(A)} = \begin{vmatrix} -(\nu_{98}+\nu_{9\ 10}+\eta_9) & \nu_{98} & \nu_{9\ 10} & \eta_9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -(\nu_{23}+\nu_{21}) & \nu_{23} & \nu_{21} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

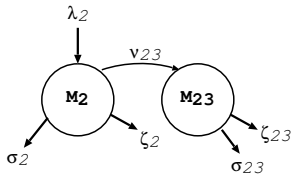
$$\mathbf{D}^{(A)} = diag\begin{vmatrix} \sigma_9+\zeta_9 & \sigma_{98}+\zeta_{98} & \sigma_{9\ 10}+\zeta_{9\ 10} & \sigma_2+\zeta_2 & \sigma_{21}+\zeta_{21} & \sigma_{23}+\zeta_{23} \end{vmatrix}$$

$$\mathbf{b}^{(A)} = \begin{vmatrix} \lambda_9 & 0 & 0 & 0 & 0 & 0 \end{vmatrix} \tag{12}$$

Figure 6: Matrices and vector describing Agent A's model.

Losses are represented through different Loss agents. An example is depicted in Fig. 8, where the agent models the failure of data transfer requests to PoA_2. This agent is characterised by a single state and its behaviour is represented only by the birth vector $\mathbf{b}$:

$$\mathbf{b} = \left| \left(n_2^{(A)} + n_2^{(B)} + \ldots + \sum_c \sum_y n_{y2}^{(c)} \right) \zeta_2 \right| \tag{13}$$



$$\mathbf{Q}^{(B)} = \begin{vmatrix} -\nu_{23} & \nu_{23} \\ 0 & 0 \end{vmatrix}$$

$$\mathbf{D}^{(B)} = \begin{vmatrix} \sigma_2 + \zeta_2 & 0 \\ 0 & \sigma_{23} + \zeta_{23} \end{vmatrix}$$

$$\mathbf{b}^{(B)} = \begin{vmatrix} \lambda_2 & 0 \end{vmatrix}$$

Figure 7: Agent B's model and matrices.

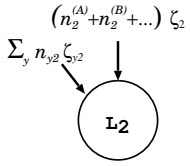


Figure 8: Loss model for PoA_2.

where $n_2^{(A)} + n_2^{(B)} + \ldots$ represent the classes of users accessing directly PoA_2, and the term $\sum_c \sum_y n_{y2}^{(c)}$ accounts for the users that switched to PoA_2 from neighbouring cells.

To solve the equations in (11), we use the *Runge-Kutta with adaptive step-size control* discretisation method [45], since the system is generally stable and this technique provides a good tradeoff between accuracy and computation time.

We remark that the model and the solution technique we developed allow us to evaluate the performance of large-scale network systems and to compute in a fast, yet accurate manner, several, relevant metrics.

5.1. Computation of performance metrics

From the evolution vector $\boldsymbol{n}^{(i)}$, which accounts for the number of agents in a given state in the coverage area $\mathcal{A}_i$, we can derive several indices that can be used to assess the system performance.

The transition rates ζ_x express the frequencies at which requests fail due to limited resources in each area, and the value of the evolution vector for the Loss agents accounts for the total losses at the corresponding PoA.

The number of requests enqueued for transmission in each PoA_x can be computed as,

$$n_x + \sum_{y \in \mathcal{X}_{\S}} n_y \,. \tag{14}$$

Similarly, we can compute the throughput associated with the generic PoA_i, denoted by $X^{(i)}$. Specifically, taking into account the speed at which each request is served as well as the number of requests being served, we can write the throughput as:

$$\boldsymbol{n}_i \sigma_i + \sum_{j \in \mathcal{X}_{\wr}} \boldsymbol{n}_{ji} \sigma_{ji} \,. \tag{15}$$

5.2. Complexity

In a nutshell, we conveniently represent the user-level dynamics through the agent-based model, and the network-level dynamics through the PoA Markov model. In particular, given the following definition:

$$N_S = \sum_c st_c \tag{16}$$

the model *complexity depends on the sum of the number of states of each PoA (N_S), as opposite to conventional techniques* resorting to full-state space generation with a complexity proportional to the product of the number of states of each PoA. Let us call N_{neigh} the average number of states belonging to neighbor cells used to determine the transition rates of the single state; we have: $N_{neigh} \leq N_S$. The complexity of each time-step used in the Runge-Kutta technique used to solve the model is $O(N_{neigh} \cdot N_S)$. The spatial complexity is instead $O(N_S)$. Considering then that the Runge-Kutta procedure is linear in the selected time horizon, the technique allows us to analyse pretty large scenarios that could not be addressed using conventional technique such as discrete event simulation.

6. Validation

We validated the model against simulation results obtained through the ns3 simulator [46]. The following sections show the network topology and the instance of the general model we considered for validating our approach. Finally, simulation and model results are compared.

6.1. Simulation Scenario

Since our model can capture different technologies and user dynamics, we select a network scenario where macrocells use the LTE technology and microcells use WiFi. Also, for sake of simplicity, we consider only vertical handovers, as illustrated by the flow chart in Fig. 9. In particular, users covered by both technologies select WiFi first, and they perform a vertical handover toward the local LTE base station (hereinafter referred to as eNB) if the throughput experienced through WiFi is lower than a given threshold. A connection is lost if the user cannot perform the handover; similarly, if the currently experienced throughput is too low, a data transfer failure occurs.

The simulated network topology, depicted in Fig. 10, consists of 5 eNBs, and 4 WiFi APs per LTE cell, with the distance between two neighbouring eNBs being set to $250\,\mathrm{m}$. The parameter setting for each technology is reported in Table 2. Traffic is generated by Constant Bit Rate (CBR) sources, with rate equal to 0.05 flow/s in cells $1, 2, 4, 5$ and to 12 flow/s in cell 3. Each flow corresponds to a $5\,\mathrm{MB}$-data transfer that needs to be delivered to a user; the size of each data packet is set to $1024\,\mathrm{B}$. The simulation results in terms of throughput, offered traffic load, and data losses are averaged over 10 runs.

Fig. 11 compares analytical and simulation results, in terms of throughput obtained via LTE and via WiFi in the overall system. Importantly, the match between analysis and the average value obtained through simulation is excellent, with a confidence level

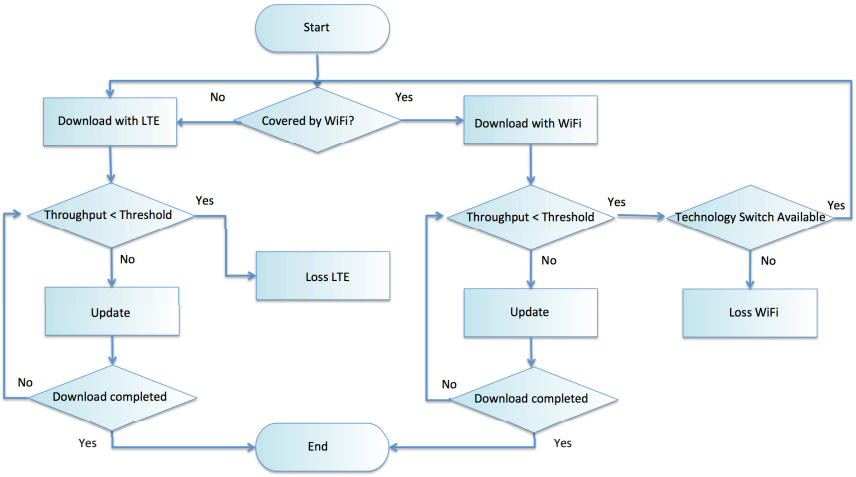


Figure 9: Technology selection flow chart used in our simulations.

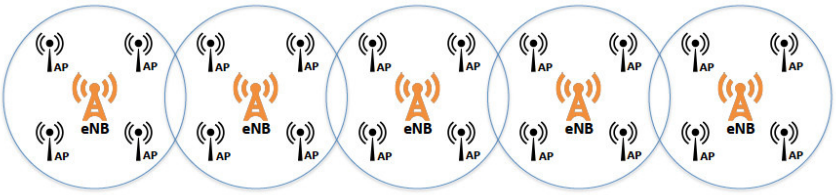


Figure 10: Topology of the scenario used for validation.

Table 2: Simulation settings: LTE, WiFi, and general parameters

LTE Parameter	Value
Carrier frequency	2.6 GHZ
Bandwidth	10 MHz (50 RB)
Distance between eNB	250 m
Height eNB	25 m
Transmission Power eNB	46 dBm
Transmission power user	24 dB
Scheduler - HARQ	Proportional Fair (PF) Scheduler
Propagation Loss Model	TwoRayGroundPropagationLossModel
WiFi Parameter	**Value**
Height AP	2.5 m
Transmission Power	15 dB
Gain in rx	1 dB
Gain in tx	1 dB
Propagation Loss model	FriisPropagationLossModel
General Parameter	**Value**
Total area	0.39 km^2
Throughput threshold	0.2 Mb/s
Request Size	1 Mbyte
Number of users	380 (300 in Cell 3 and 20 in other cells)
User antenna height	1.5 m
Number of cells	5
Number of AP per cell	4
Traffic	UDP
MTU	1500 byte
eNb buffer size	100 Mb

of 90%. Due to the presence of CBR traffic sources and stationary system conditions, the throughput results to be almost constant during the simulation.

Note that the simulation model implements all the TCP/IP communication levels; the simulation configuration parameters are also reported in Table 2. Instead, our analytical model abstracts most of the layers of the communication stack. In particular, it takes into the account: the resource access for each of the considered technologies, the data transmissions, the handovers, as well as a connection drop due to lack of resources or a user leaving the system. The flow chart in Fig. 9, which exactly describes the technology selection strategy, is implemented in both the simulation and the general model instance we derived for the considered scenario. We also remark that the system scenario under study is configured in simulation through the ns3 APIs, while in the proposed analytical model it is mapped onto the functions used in the matrix elements appearing in (11). In the following section, we exploit our model to investigate the system dynamics, considering a real-world network topology and user behaviour.

7. Exploitation

We now focus on a real-world scenario, obtained by considering the data available in [47] from which we selected both the user mobility and the traffic demand traces. Such traces refer to 13 consecutive hours in a single day at Disney World, Orlando, in the USA; however, for readability of the results, we limit our study to a three-hour period showing the traffic load time evolution, from opening time of the park to the time of the first parade (peak hour). The scenario represents a crowded area in which, using the information available in [48], we identified the exact location of the LTE

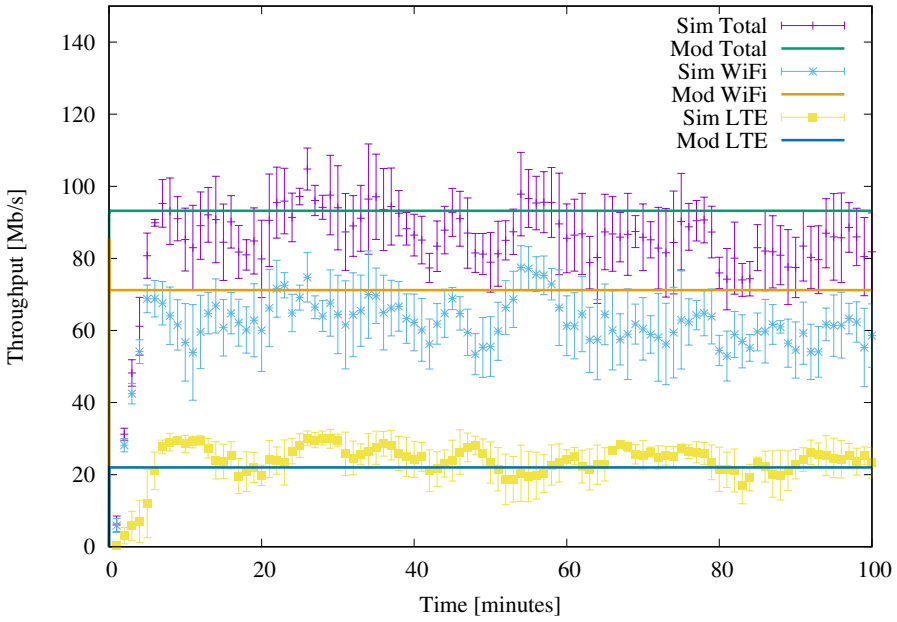


Figure 11: Model validation: comparison between analytical and simulation results in terms of throughput time evolution. Different markers correspond to WiFi (star), LTE (square), and total (dash); for simulation results, we report also the confidence interval (90%).

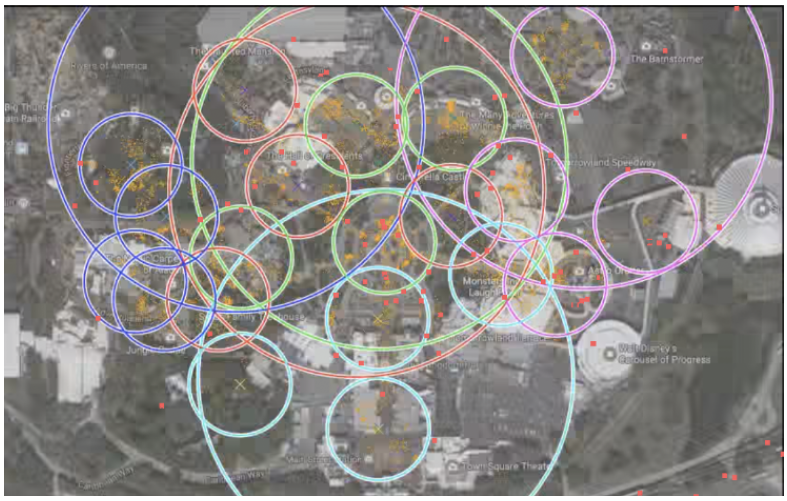


Figure 12: Real-world scenario with antennas and PoA coverages.

eNBs. The considered area is reported in Fig. 12 where there are 5 eNBs. In each cell, we placed 4 WiFi APs by applying to the aforementioned traces the k-means technique so as to identify the areas with higher user density (in the figure, such areas are represented by smaller circles of the same color as that of the corresponding eNB). We remark that, so doing, the AP locations turn out to be quite similar to those of commercials hotspots (the latter being obtained according to [49] and highlighted by red dots in the figure). Users are assigned to WiFi APs and to LTE eNBs according to their position; a snapshot of such assignment is illustrated in Fig. 13. Although our technique could support arbitrarily shaped communication ranges (as representative of complex transmission scenarios), we have resorted to circular areas, with different radius depending on the considered technology. Specifically, LTE eNBs range is set to 150 m, according to the real power parameter available on [48]; the AP transmission radius, instead, is set to 40 m, so as to account for the high user density.

Such a scenario and system model are then used to determine the user agents and their corresponding classes (see Section 4) as time varies, according to the aforementioned mobility trace. This in turn determines the arrival rate of data downloading requests at each PoA. Note that, although in this work we derived the location of the users directly from the data traffic traces, our technique can be applied in the presence of any model describing the users distribution. Overall, the system we obtained perfectly matches the scenario presented in Fig. 1, and the users can be described by the agent shown in Fig. 3. Finally, as mentioned in Section 6, we only consider vertical handovers.

From the traces and available consumers statistics that indicate around 10 million visitors per year, we considered a population of $32,800$ users. The above scenario is thus representative of large-scale network systems, for which our model is particularly useful; indeed, it allows to evaluate the average behaviour without requiring cumbersome simulations representing each user independently.

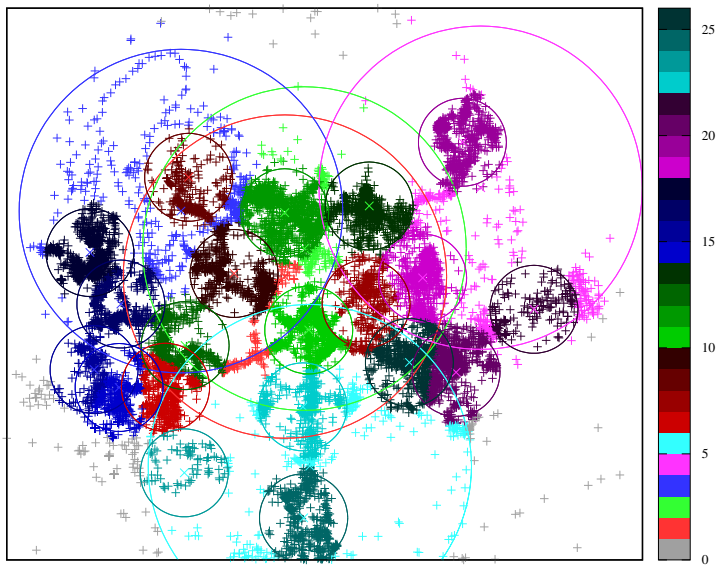


Figure 13: Users partitioning into coverage areas.

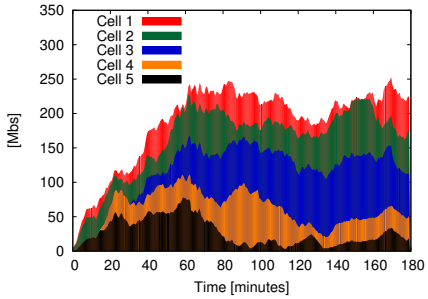
7.1. Network evaluation

To evaluate the network performance, we present a number of metrics obtained through our analytical model, namely, throughput, traffic load at the PoAs, data losses, and throughput fairness. Specifically: (5) and (15) determine the throughput; (14), providing the number of users connected to a specific cell, allows us to compute the traffic load at each PoAs; (9) and (10), which account for users that have lost their connection, allow for the computation of the data losses; finally, fairness in terms of user throughput is introduced in this section and is directly obtained from the model solution in (11).

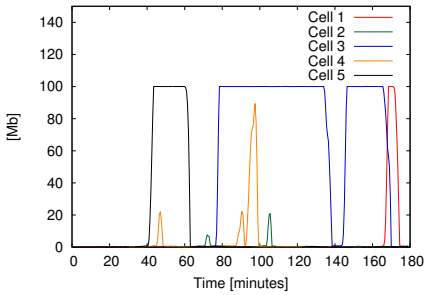
We recall that the analysis focuses on a three-hour period, from the park opening to the first peak user density. Results referring to the five cells are shown in Fig. 14: during the 3-hour period, the zone covered by Cell 3 is the most visited, leading to a high user, hence data traffic, density (see Fig. 14a). The consequence of such traffic distribution in terms of traffic load at the eNBs can also be observed in Fig. 14b.

Cell 5 receives the largest number of visits in the first hour, and, when the APs therein cannot provide a sufficiently high throughput any longer, several vertical handovers occur, thus increasing the traffic load at the eNB and eventually causing traffic losses (Fig. 14c). A similar behavior can be observed for Cell 3 where, however, traffic congestion last longer, thus leading to higher losses. Finally, toward the end of the considered time interval, also Cell 1 experiences some losses due to a peak in the user density. No losses instead are recorded for Cells 2 and 4, always exhibiting a low traffic load.

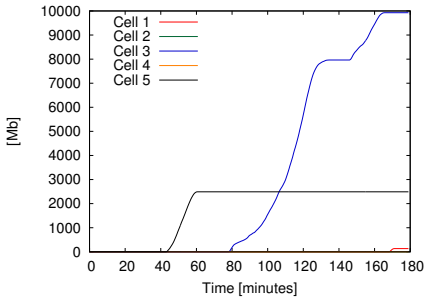
Fig. 14d shows the throughput fairness computed through Jain's index [50], in order to show the level of fairness in the resource allocation across the users in each LTE



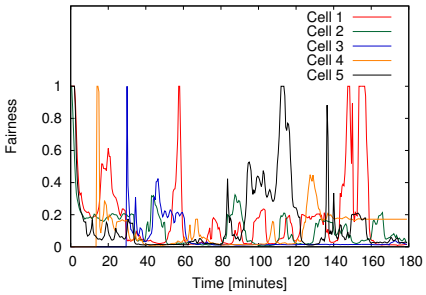
(a) Throughput distribution across the LTE cells



(b) Traffic load at the eNBs, with buffer size set to 100 Mb



(c) Data losses



(d) User fairness in terms of throughput

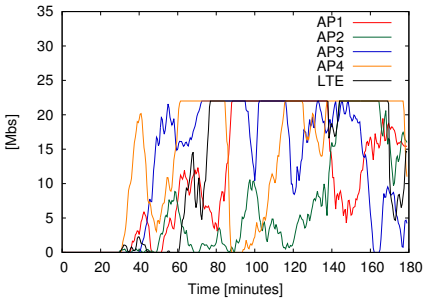
Figure 14: Performance metrics related to LTE cells.

cell. We recall that the closer to 1 the values of the Jain's index, the higher the level of fairness of the network service. Within each cell, the index varies depending on the users' spatial distribution, but it can be noted that, in Cell 3, it drops significantly as the cell gets more crowded.

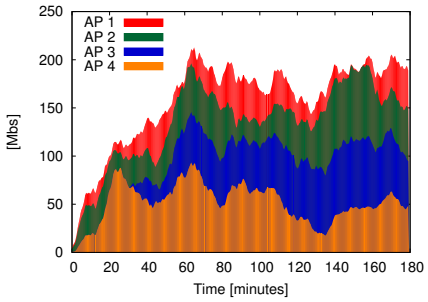
The results presented above suggest that Cell 3 is the most critical one; thus, in the following we focus on this cell in order to compare the WiFi and LTE technologies, as well as the effect of vertical handovers.

Looking at the plots in Fig. 15, we notice that, within Cell 3, the first AP where the throughput drops below the threshold is AP4 (see Fig. 15a and 15b around the 60th minute). The vertical handovers from WiFi to LTE cause an increase in the eNB traffic load; when also AP 3 becomes congested (around the 75th minute) and more traffic flows are moved towards LTE, the eNB becomes saturated and data traffic starts to be dropped, as shown in Fig. 15c. This behaviour is observed approximately till the 175th minute, when the rush hour period ends and no losses occur any longer.

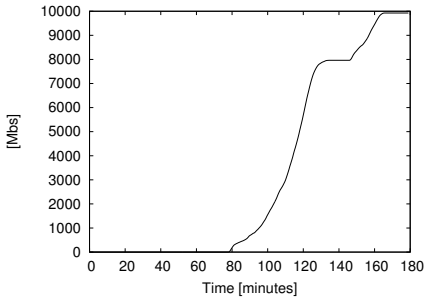
Fig. 15d depicts the level of fairness in terms of throughput, for each PoA of the cell. It can be observed that the WiFi APs can provide better fairness than the LTE eNB: indeed, the access point coverage area is smaller than the one of LTE and the WiFi users experience more similar propagation conditions, hence throughput performance.



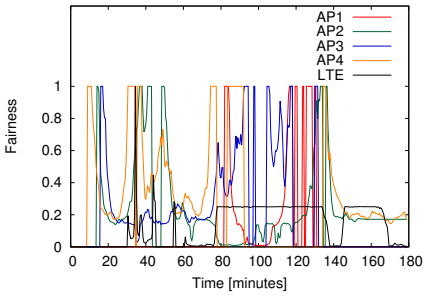
(a) Throughput of WiFi APs and LTE eNBs



(b) Throughput distribution across the APs



(c) LTE data losses



(d) User fairness in terms of throughput: comparison between WiFi and LTE

Figure 15: Performance related to Cell 3, and comparison between WiFi and LTE.

8. Conclusions

Given the high level of heterogeneity and densification of new-generation mobile networks, we developed a flexible, yet highly accurate, analytical model for the performance study of such complex systems. We built an analytical framework based on the Markovian agent formalism, which well captures all main aspects of a heterogeneous, multi-tier network. By solving the model through a mean-field approach, we derived several important performance metrics and showed that our model closely mimics the system behavior in the case of high user density. We then validated our analysis by simulation and investigated the performance of the network system in a real-world, large-scale scenario. The results demonstrate how the model can be conveniently used to allocate resources in a heterogeneous network and across the different tiers.

We remark that, by capturing both the technology diversity and the multiplicity of communication resources that characterize 5G-and-beyond network systems, our model can provide useful guidelines for an efficient and effective management of communication resources. Furthermore, being highly scalable, it enables the analysis of a very high number of simultaneous connections – a key feature of several 5G use cases.

References

[1] Mobile data traffic outlook, https://www.ericsson.com/en/mobility-report/reports/june-2019/mobile-data-traffic-outlook, (Accessed Accessed in Oct. 2019).

[2] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, J. Rodriguez, Millimeter-wave massive mimo communication for future wireless systems: A survey, IEEE Communications Surveys Tutorials 20 (2) (2018) 836–869.

[3] Q. Zhu, X. Wang, Z. Qian, Energy-efficient small cell cooperation in ultra-dense heterogeneous networks, IEEE Communications Letters 23 (9) (2019) 1648–1651.

[4] T. O. Olwal, K. Djouani, A. M. Kurien, A survey of resource management toward 5g radio access networks, IEEE Communications Surveys Tutorials 18 (3) (2016) 1656–1686.

[5] R. Vannithamby, S. Talwar, Distributed Resource Allocation in 5G Cellular Networks, Wiley, 2017. doi:10.1002/9781118979846.ch8.
URL https://ieeexplore.ieee.org/document/8045143

[6] S. Chen, J. Zhao, M. Ai, D. Liu, Y. Peng, Virtual rats and a flexible and tailored radio access network evolving to 5g, IEEE Communications Magazine 53 (6) (2015) 52–58.

[7] M. Benaim, J.-Y. L. Boudec, A class of mean field interaction models for computer and communication systems, Performance Evaluation 65(11-12) (2008) 823–838.

[8] A. Bobbio, M. Gribaudo, M. Telek, Analysis of large scale interacting systems by mean field method, in: 5th International Conference on Quantitative Evaluation of Systems - QEST2008, St. Malo, 2008.

[9] M. Gribaudo, D. Cerotti, A. Bobbio, Analysis of on-off policies in sensor networks using interacting markovian agents, in: 4th International Workshop on Sensor Networks and Systems for Pervasive Computing - PerSens 2008, Hong Kong, 2008.

[10] F. Cordero, D. Manini, M. Gribaudo, Modeling biological pathways: an object-oriented like methodology based on mean field analysis, in: the Third International Conference on Advanced Engineering Computing and Applications in Sciences(ADVCOM), IEEE Computer Society Press, 2009, pp. 193–211.

[11] Y. Khadraoui, X. Lagrange, A. Gravey, A survey of available features for mobile traffic offload, in: European Wireless 2014; 20th European Wireless Conference, 2014, pp. 1–4.

[12] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. Wong, R. Schober, L. Hanzo, User association in 5g networks: A survey and an outlook, IEEE Communications Surveys Tutorials 18 (2) (2016) 1018–1044.

[13] X. Ge, X. Li, H. Jin, J. Cheng, V. C. M. Leung, Joint user association and user scheduling for load balancing in heterogeneous networks, IEEE Transactions on Wireless Communications 17 (5) (2018) 3211–3225.

[14] F. Pervez, M. Jaber, J. Qadir, S. Younis, M. A. Imran, Memory-based user-centric backhaul-aware user cell association scheme, IEEE Access 6 (2018) 39595–39605.

[15] M. M. Hasan, S. Kwon, S. Oh, Frequent-handover mitigation in ultra-dense heterogeneous networks, IEEE Transactions on Vehicular Technology 68 (1) (2019) 1035–1040.

[16] E. Demarchou, C. Psomas, I. Krikidis, Mobility management in ultra-dense networks: Handover skipping techniques, IEEE Access 6 (2018) 11921–11930.

[17] IEEE Standard for Local and Metropolitan Area Networks−Part 21: Media Independent Handover, IEEE Std. 802.21-2008.

[18] E. Aryafar, M. W. A. Keshavarz-Haddad, M. Chiang, RAT selection games in HetNets, IEEE INFOCOM (2013) 998–1006.

[19] K. Zhu, E. Hossain, D. Niyato, Pricing, spectrum sharing, and service selection in two-tier small cell networks: A hierarchical dynamic game approach, IEEE Transactions on Mobile Computing 13 (8).

[20] A. A. Alabdel Abass, R. Kumbhkar, N. B. Mandayam, Z. Gajic, Wifi/lte-u coexistence: An evolutionary game approach, IEEE Transactions on Cognitive Communications and Networking 5 (1) (2019) 44–58.

[21] H. Liu, C. Maciocco, V. Kesavan, A. Low, Energy efficient network selection and seamless handovers in mixed networks, in: IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks Workshops, 2009, pp. 1–9.

[22] H. Petander, Energy-aware network selection using traffic estimation, ACM workshop on Mobile internet through cellular networks (2009) 55–60.

[23] P. Gawlowicz, A. Zubow, A. Wolisz, Enabling cross-technology communication between lte unlicensed and wifi, in: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 144–152.

[24] A. Garcia-Saavedra, P. Patras, V. Valls, X. Costa-Perez, D. J. Leith, Orla/olaa: Orthogonal coexistence of laa and wifi in unlicensed spectrum, IEEE/ACM Transactions on Networking 26 (6) (2018) 2665–2678.

[25] A. Roy, P. Chaporkar, A. Karandikar, Optimal radio access technology selection algorithm for lte-wifi network, IEEE Transactions on Vehicular Technology 67 (7) (2018) 6446–6460.

[26] M. Gerasimenko, N. Himayat, S.-P. Yeh, S. Talwar, S. Andreev, Y. Koucheryavy, Characterizing performance of load-aware network selection in multi-radio (wifi/lte) heterogeneous networks, in: IEEE Globecom Workshops (GC Wkshps), 2013, pp. 397–402.

[27] T. Wan, P. Li, X. Wang, Y. Wang, T. Guo, Y. Cao, A comprehensive survey on mobile data offloading in heterogeneous network, Wireless Networks 25 (2) (2019) 573–584.

[28] M. Morant, R. Llorente, Performance analysis of multiple radio-access provision in a multicore-fibre optical fronthaul, Optics Communications 436 (2019) 161–167.

[29] T. NAKAMURA, A. BENJEBBOUR, Y. KISHIYAMA, S. SUYAMA, T. IMAI, 5G radio access: Requirements, concept and experimental trials, IEICE Transactions on Communications E98-B (8) (2015) 1397–1406.

[30] R. Odarchenko, J. Gimenez, Y. Sulema, B. Altman, S. Petersen, Multilink solution for 5g: Efficiency experimental studies, in: 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), 2019, pp. 336–339.

[31] P. Gawłowicz, A. Zubow, Demo abstract: Practical cross-technology radio resource management between lte-u and wifi, in: IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2018, pp. 1–2.

[32] M. Simsek, T. Hößler, E. Jorswieck, H. Klessig, G. Fettweis, Multiconnectivity in multicellular, multiuser systems: A matching- based approach, Proceedings of the IEEE 107 (2) (2019) 394–413.

[33] M. Yao, M. Sohul, V. Marojevic, J. H. Reed, Artificial intelligence defined 5g radio access networks, IEEE Communications Magazine 57 (3) (2019) 14–20.

[34] M. G. Kibria, K. Nguyen, G. P. Villardi, W. Liao, K. Ishizu, F. Kojima, A stochastic geometry analysis of multiconnectivity in heterogeneous wireless networks, IEEE Transactions on Vehicular Technology 67 (10) (2018) 9734–9746.

[35] D. Bruneo, S. Distefano, F. Longo, A. Puliafito, M. Scarpa, Evaluating wireless sensor node longevity through markovian techniques, Computer Networks 56 (2) (2012) 521 – 532. doi:https://doi.org/10.1016/j.comnet.2011.10.003.
URL http://www.sciencedirect.com/science/article/pii/S1389128611003690

[36] Y. Ren, T. Phung-Duc, Y. Liu, J. Chen, Y. Lin, ASA: adaptive VNF scaling algorithm for 5g mobile networks, in: 7th IEEE International Conference on Cloud Networking, CloudNet 2018, Tokyo, Japan, October 22-24, 2018, IEEE, 2018, pp. 1–4. doi:10.1109/CloudNet.2018.8549542.
URL https://doi.org/10.1109/CloudNet.2018.8549542

[37] N. Gast, B. V. Houdt, A refined mean field approximation, in: K. Psounis, A. Akella, A. Wierman (Eds.), Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS 2018, Irvine, CA, USA, June 18-22, 2018, ACM, 2018, p. 113. doi:10.1145/3219617.3219663.
URL https://doi.org/10.1145/3219617.3219663

[38] D. Bruneo, M. Scarpa, A. Bobbio, D. Cerotti, M. Gribaudo, Markovian agent modeling swarm intelligence algorithms in wireless sensor networks, Perform. Eval. 69 (3-4) (2012) 135–149. doi:10.1016/j.peva.2010.11.007.
URL https://doi.org/10.1016/j.peva.2010.11.007

[39] M. C. Guenther, J. T. Bradley, Mean-field analysis of data flows in wireless sensor networks, in: Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering, ICPE '13, ACM, New York, NY, USA, 2013, pp. 51–62. doi:10.1145/2479871.2479882.
URL http://doi.acm.org/10.1145/2479871.2479882

[40] M. Gribaudo, D. Manini, C.-F. Chiasserini, Studying mobile internet technologies with agent based mean-field models., in: ASMTA, 2013, pp. 112–126.

[41] C. Chiasserini, M. Gribaudo, D. Manini, Traffic offloading/onloading in multi-rat cellular networks, in: Proceedings of the IFIP Wireless Days, WD 2013, Valencia, Spain, November 13-15, 2013, 2013, pp. 1–7. doi:10.1109/WD.2013.6686526.
URL https://doi.org/10.1109/WD.2013.6686526

[42] B. Yang, W. Guo, Y. Jin, S. Wang, Smartphone data usage: downlink and uplink asymmetry, Electronics Letters 52 (3) (2016) 243–245.

[43] T. G. Kurtz, Solutions of ordinary differential equations as limits of pure jump markov processes, Journal of Applied Probability 7 (1) (1970) 49–58.
URL http://www.jstor.org/stable/3212147

[44] M. Heusse, F. Rousseau, G. Berger-Sabbatel, A. Duda, Performance anomaly of 802.11b, in: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, Vol. 2, 2003, pp. 836 – 843 vol.2.

[45] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd Edition, Cambridge University Press, New York, NY, USA, 2007.

[46] Ns-3 network simulator, https://www.nsnam.org/releases/ns-3-29/documentation/ (2011-18).

[47] Crawdad - a community resource for archiving wireless data at dartmouth mirror provided by the school of computer science at the university of st andrews, https://uk.crawdad.org/.

[48] Cell mapper, https://www.cellmapper.net/map.

[49] wigle, https://www.wigle.net.

[50] R. Jain, D. Chiu, W. Hawe, A quantitative measure of fairness and discrimination for resource allocation in shared systems, Digital Equipment Corporation,Technical Report DEC-TR-301, Tech. Rep., technical Report (1984).