

Traffic Offloading/Onloading in Multi-RAT Cellular Networks

Carla-Fabiana Chiasserini
Politecnico di Torino
Torino, Italy
Email: chiasserini@polito.it

Marco Gribaudo
Politecnico di Milano
Milano, Italy
Email: gribaudo@elet.polimi.it

Daniele Manini
Università di Torino
Torino, Italy
Email: manini@di.unito.it

Abstract—We analyze next generation cellular networks, offering connectivity to mobile users through multiple radio access technologies (RATs), namely LTE and WiFi. We develop a framework based on the Markovian agent formalism, which can model several aspects of the system, including user traffic dynamics and radio resource allocation. In particular, through a mean-field solution, we show the ability of our framework to capture the system behavior in flash-crowd scenarios, i.e., when a burst of traffic requests takes place in some parts of the network service area. We consider a distributed strategy for the user RAT selection, which aims at ensuring high user throughput, and investigate its performance under different resource allocation schemes.

I. INTRODUCTION

Recent results highlight that one of the most challenging issues in the field of communication networks today is coping with the exponential growth of wireless data traffic. The average smartphone is expected to generate 2.6 GB of traffic per month by 2016, with a global mobile data traffic that will increase 18-fold by that time [4]. To increase the capacity of cellular networks and, thus, accommodate such high data-traffic loads, the new Long-Term Evolution (LTE) technology has been introduced, which is deemed to achieve a spectrum efficiency as high as 1.75 bits/s/Hz, a downlink peak throughput of 100 Mb/s and a latency of few milliseconds.

The fast uptake of mobile data services, however, indicates that these solutions are not sufficient to meet the intense user demand in many high-density settings. Thus, a new trend, usually referred to as *mobile data offloading*, has emerged. That is, while the cellular infrastructure will continue to provide high levels of quality of service (QoS), besides the support for high-mobility users, it will be complemented with alternative wireless access technologies such as WiFi hotspots. Data traffic should therefore be offloaded whenever possible towards such hotspots, at the price of a possible degradation in the QoS experienced by the users [10], [15].

Such scenario calls for a new access network architecture, whose distinctive feature is the availability of multiple radio access technologies (RATs). According to this paradigm, the network is composed of base stations (BSs) equipped with more than one radio interface (namely, LTE and WiFi), through which users can access the Internet [2], [5]. Examples of multi-RAT BS commercial products can be already found on the market, e.g., [1], [2]. Furthermore, several proposals

have appeared in the literature, presenting vertical handover mechanisms or schemes for letting the users always connect to the best network access point, see, e.g., [8], [11], [20] and references therein.

However, most of these solutions are based on the optimistic assumption that users can become aware of the throughput they will receive when using a RAT. What is still to be investigated is a fully-distributed mechanism to make the above access network work efficiently, in absence of a-priori information on the throughput to be expected. Additionally, we would need a framework that, with low complexity, allows the evaluation and comparison of the different proposals.

In this paper, we address the above shortcomings and develop a framework for the analysis of strategies for handling the Internet connectivity of mobile users. We consider that the network service area is covered by a number of BSs, each of them hosting both an LTE and a WiFi radio interface. A user can connect to the Internet through either technologies, provided that she is under coverage and enough radio resources are available. We then focus on a specific, fully-distributed strategy for RAT selection. The strategy takes the user's viewpoint and aims at offering the QoS desired by the user at the best price.

As already done by widely popular communication devices (e.g., smartphones), our strategy first lets users connect to a WiFi hotspot whenever available. The rationale behind this is that WiFi connectivity is much less costly than the cellular one: through WiFi the per-byte cost of data transfers can be reduced by 70% per one estimate [3]. However, users may not receive a sufficiently high data throughput by connecting to WiFi, due to, e.g., bad radio propagation conditions or a high congestion level at the hotspot [5], [14], [18]. In this case, a service upgrade can be performed by onloading their data traffic to the LTE network if cellular radio resources are available. An offloading decision can be taken again if LTE becomes overloaded, or if the WiFi network can offer sufficiently high throughput.

To analyze the above system and efficiently allocate the radio resources, we propose an analytical model, which leverages mean-field analysis [6], [7]. Indeed, unlike other techniques, such as Queueing Networks, Stochastic Petri Net, or Process Algebras, mean-field analysis allows us to account for the spatial distribution of the communication nodes (BSs

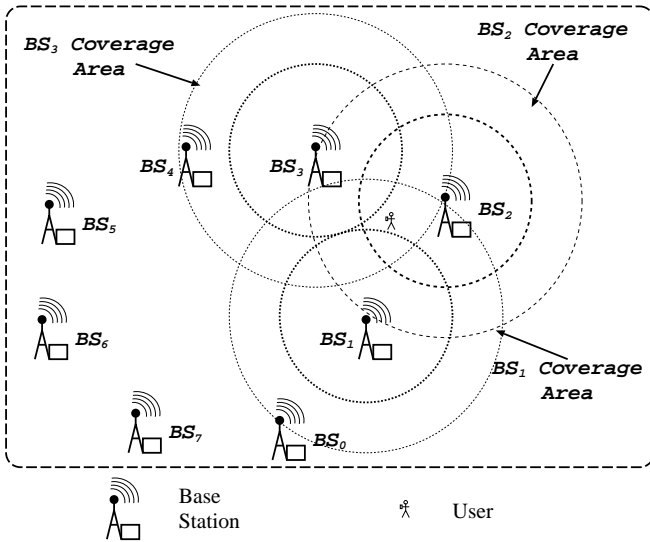


Fig. 1. An example of the topology under study.

and users) in the system. This is clearly of fundamental importance, as a user can access a BS only if its position is within the coverage area of that BS. The model is then solved by resorting to a method based on the Markovian Agent formalism [12], and exploiting the results in [9]. We remark that our analytical framework could be extended to investigate other RAT selection strategies as well. Using the above approach, we analyze the system performance and the level of QoS that the users experience when different resource allocations techniques are implemented.

The rest of the paper is organized as follows. In Section II, we introduce the system under study, along with the distributed RAT selection strategy that we consider. The analytical model we develop and the model solution are presented in Sections III and IV, respectively. In Section V, we describe the resource allocation policies that we assume in our performance evaluation and show the results that we obtain. Finally, in Section VI we draw our conclusions and provide directions for future research.

II. SYSTEM SCENARIO

Network scenario. We consider a urban area (typically characterized by high user density), covered by N LTE BSs. We refer to the coverage area of the generic BS_i as L_i , with $i = 0, \dots, N-1$. Co-located with the LTE interface, there is a WiFi radio (IEEE 802.11a/g/n), so as to implement a hotspot whose coverage is assumed for simplicity to coincide with that provided by the LTE technology (e.g., a coverage range of 100-200 m). Coverage areas of neighboring BSs may overlap. We denote by $neigh(i)$ the set of BSs that are neighbors of BS_i . Users located in the area $L_i \cap L_j$ can access either BS_i or BS_j . A graphical representation of the the network scenario is depicted in Fig. 1, where $N = 6$ and, e.g., $neigh(2) = \{3, 1\}$.

We only consider data transfers, such as content downloading or video streaming, since voice traffic is not currently supported by real-world LTE networks. Also, we focus on

downlink transfers (from the BS to the users) since traffic is typically asymmetric, with a large amount of data flowing from the Internet towards the users.

We denote by $n_{WiFi\ resources}^{[i]}$ the number of frequency channels that are available for WiFi communications at BS_i and assume that, given two BSs with overlapping coverage, they use different channels so as to avoid interference between simultaneous transmissions. As for LTE, we consider that a frequency division duplex (FDD) technique is used for data transmission, as typically done in practical systems. We set the central operating frequency to 2.6 GHz and the channel bandwidth to 10 MHz. We remark that we set these values for concreteness, but any other value allowed by the LTE technology could be considered as well [17]. Time is then divided into frames that are 10 ms long; every frame is further divided into 10 subframes of 1 ms each. As specified by 3GPP, a frequency channel is divided into several narrow-band subchannels, which are grouped in disjoint subsets of 12 subchannels each. The usage of a subset of 12 subchannels for a 1-ms duration (i.e., a subframe) represents a Physical Radio Block (PRB). We refer to it as the resource unit in LTE; thus, given a 10-MHz bandwidth, $n_{LTE\ resources}^{[i]} = 50$ LTE resources are available [17] at BS_i . Note that the BS may allocate one or more resource units to transmit at the same time towards the same user. As for the interference, similarly to what done for WiFi, we consider that an LTE resource unit can be used by different BSs at the same time only if their coverage areas do not overlap.

As mentioned, users that are under the coverage of more than one BS, can connect to any of them, although with different QoS. In particular, for each technology (either LTE or WiFi), it is fair to assume that the link quality decreases as the distance between a user and a BS increases, and the better the link quality, the higher the transmission rate that the link end points can use. As an example, in Fig. 1, the inner circle within each coverage area represents the zone where the signal of the corresponding BS is stronger. Finally, we consider that the users are pedestrian and do not significantly move while receiving a data transfer from the Internet, as the latter is expected to last just few seconds.

RAT selection strategy. The strategy we envision aims at ensuring that, for each traffic flow she starts, a user selects the technology which provides her with sufficiently high throughput while minimizing the access cost. The strategy is fully distributed, i.e., users independently decide which RAT to use in order to download a content from the Internet and they cannot rely on detailed, a priori information on the throughput they will experience when connected to an access interface.

More specifically, a user lists the available WiFi and LTE interfaces, according to the received signal strength (RSS). Then, she first connects to the WiFi interface located at the BS from which she receives the strongest signal (hereinafter referred to as *local* interface or BS). If the obtained throughput is below a given threshold, the user switches her traffic flow to the WiFi interface that is second in the list. We call such interface (and the corresponding BS) *neighboring* interface

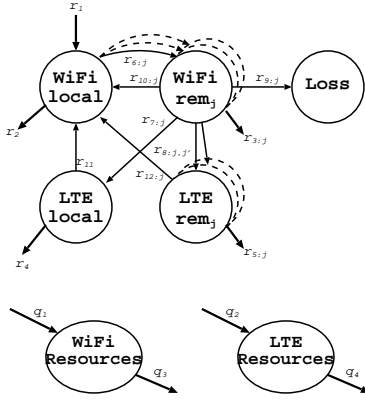


Fig. 2. Mean-field models of a user agent (left) and of a BS agent (right).

(BS). Depending on the experienced throughput, the user migrates again if not satisfied. This time, she unloads her traffic to the LTE interface of the local BS. As the last option, she moves to the LTE interface of a neighboring BS. Furthermore, in order to keep the Internet connectivity cost as low as possible, a user currently downloading data from an LTE interface periodically will try to offload her traffic to her local WiFi. Similarly, a user receiving data from a WiFi neighboring interface periodically will try to migrate to the local one.

III. NETWORK MODEL

We call *agent* a portion of the model describing the behavior of the communication nodes within a given coverage area L_i . We use three agents per coverage area so as to describe the state of (1) the network connectivity of users located within L_i , (2) the WiFi radio resources at BS_i , and (3) the LTE radio resources at BS_i . We also denote by AG_i the *access group*: the set of users that access the network using BS_i .

We use the Markovian Agent formalism [12] to model the system, and we resort to the solution techniques outlined in [9] to analyze it. In particular, Markovian agents are a special type of Markov chains in which each state is characterized by the number of agents that are in that specific state in the considered location, and where the transition rate among states can depend on the number of agents in given states of the neighboring agents. The agents used to describe the coverage areas of the considered system are shown in Fig. 2. The upper part corresponds to the Markovian agent used to model the user behavior (hereinafter referred to as user agent), while the two lower agents correspond to the WiFi and LTE resources of the BS.

Each user agent has two states that represent whether the network is accessed using the WiFi or LTE resources of BS_i (called, respectively, *WiFi local* and *LTE local*). Then, there is a set of states that represent the users accessing the network using the resources of a neighboring BS. In particular, the model of a generic coverage area L_i includes, for each neighboring BS_j , a state *WiFi rem_j* and a state *LTE rem_j* representing the users that access the network using, respectively, the WiFi and the LTE resources of the

neighboring BS. Finally, each user agent includes a *Loss* state accounting for all the packets that could not be transmitted due to lack of resources (at both the local and the neighboring BSs). The incoming arrows correspond to external events: in our case they model the arrival of new packets that have to be transmitted. Indeed, as mentioned earlier, new users always start by accessing local WiFi; then, they move to a different BS or interface if available resources are not sufficient. Similarly, outgoing arrows represent agents that exit the system, which, in our case, correspond to packets being correctly transmitted from the BS and through the interface that the corresponding state represents.

The state of the BS is modeled by two states, accounting for the radio resources available there: frequency channels in WiFi and PRBs in LTE. The incoming/outgoing arcs account for the addition/removal of a radio resource at the BS, whenever dynamic resource allocation policies are implemented.

A. Functional rates

In Markovian Agents, the interaction among the agents is modeled through functional transition rates that represent the speed at which agents move from one state to another. In this work, we have extended the functional rates presented in [13] in order to consider both an arbitrary number of neighboring BSs for each BS and time-dependent resource allocation policies. In the following, we denote by $n_s^{[i]}$ the number of agents in state s , related to coverage area L_i .

Arrivals. We model the service requests per second generated by users located in coverage area L_i with the functional rate $r_1^{[i]} = \lambda_i(t)$ ($i = 0, \dots, N - 1$). In this way, request generation can exhibit both a spatial and a time dependency. We will exploit this feature to generate burst of requests at specific locations, and to study the corresponding network performance.

WiFi Service. Rate $r_2^{[i]}$ refers to the usage of the local WiFi radio resources. Under the assumption that the IEEE 802.11a/g technologies are used, it can be expressed as:

$$r_2^{[i]} = \sigma_{WiFi}^{[i]} = \frac{\mu_{WiFi} \times n_{WiFi \text{ resources}}^{[i]}}{n_{WiFi \text{ local}}^{[i]} + \alpha \sum_{j \in \text{neigh}(i)} (n_{WiFi \text{ rem}_j}^{[j]})} \quad (1)$$

where: μ_{WiFi} is the WiFi throughput (connection speed) when there is only one user accessing the interface, and $\alpha \geq 1$ is a factor taking into account the throughput reduction due to far-away users in L_j ($j \in \text{neigh}(i)$) accessing the WiFi interface at BS_i . More specifically, α accounts for the anomaly effect [16] typical of the 802.11a/g technologies: “slower” transmissions occupy the channel longer preventing other (faster) transmissions from occurring. Rates $r_{3;j}^{[i]}$ account for the rate at a neighboring BS and are computed similarly to $r_2^{[i]}$: if BS_j is the neighboring BS (represented by state *WiFi rem_j*), then $r_{3;j}^{[i]} = \sigma_{WiFi}^{[j]}$.

LTE Service. Let now us define $\sigma_{LTE}^{[i]}(j)$ as the LTE service rate experienced by a user in L_j accessing BS_i . We then have $r_4^{[i]} = \sigma_{LTE}^{[i]}(i)$ and $r_{5;j}^{[i]} = \sigma_{LTE}^{[j]}(i)$. The LTE service rate can

be written as:

$$\sigma_{LTE}^{[i]}(j) = \frac{\mu_{LTE}^{i,j} \times n_{LTE \text{ resources}}^{[i]}}{n_{LTE-local}^{[i]} + \sum_{k \in \text{neigh}(i)} (n_{LTE \text{ rem}_i}^{[k]})} \quad (2)$$

where $\mu_{LTE}^{i,j}$ is the LTE throughput (connection speed) corresponding to one PRB assigned by BS_{*i*} to a user located within L_j . Indeed, each PRB is assigned exclusively to one user, but, depending on the radio propagation conditions, the user can employ a higher or a lower transmission rate, motivating the dependency of the parameter on the user location (L_j).

Switching to neighboring WiFi interfaces. We assume that a user switches to the WiFi interface of a neighboring BS if its service rate at the local WiFi drops below a minimum threshold (μ_{min}). If more than one neighboring BS is available, the one with the lowest traffic load is selected (in case of a tie, the BS is chosen randomly). The latency due to the interface switching is denoted by $1/\mu_{BS-Sw}$ and the rate at which this event occurs is indicated by $r_{6;j}^{[i]}$.

Switching from neighboring WiFi to LTE. If also the WiFi service rate in neighboring BSs drops below μ_{min} , users try to migrate to LTE by performing a technology switch. Priority is given to the local BS (rates $r_{7;j}^{[i]}$). If however LTE resources at the local BS are already saturated, a neighbor LTE is tried (rates $r_{8;j,k}^{[i]}$). Note that in this case the switch goes from WiFi in a neighboring coverage area L_j , to LTE in another neighboring area L_k . Indeed, although connectivity through LTE is more costly, we expect that users are willing to pay a higher price provided that they can obtain a sufficiently high connection speed. The user migration takes place with a latency due to the technology switch of $1/\mu_{\tau-Sw}$, and the target BS is the one with the lowest traffic load level among those that have at least a fraction γ_{min} of available resources. This avoids overloading the LTE interface and, thus, the need to move traffic flows back and forth (i.e., the so-called ping-pong effect).

Loss of requests. A user first joins the local WiFi; if saturated, it tries to join the neighboring WiFi interface that exhibits the lowest traffic load. If the experienced QoS is not good enough, the user tries to connect to the local LTE, or to the LTE of a neighboring BS as the last option. If none of the neighboring LTE interfaces can provide the user with a sufficient service level, the connection request is lost. Rate $r_{9;j}^{[i]}$ in Fig. 2 accounts for the losses that are experienced in this case.

Upgrade. If the traffic load in one of the less expensive or more efficient interfaces decreases, the user might chose to improve its service. For simplicity, we consider only upgrading to local WiFi from either neighboring WiFi (rates $r_{10;j}^{[i]}$) or LTE (rates $r_{11}^{[i]}$ and $r_{12;j}^{[i]}$) interfaces. In any of these cases, when the service rate at the local WiFi ($r_2^{[i]}$) is beyond a given threshold μ_{max} , the users migrate to the local WiFi interface (i.e., high speed connectivity at low cost). The time required by the transition depends on the type of switch that is performed: $1/\mu_{BS-Sw}$ if the user comes from a neighboring WiFi, and $1/\mu_{\tau-Sw}$ if the upgrade implies a technology switch.

Radio resources allocation. One of the goals of this work is to show how dynamically varying the LTE or WiFi radio resource allocation at BSs can improve the system performance. WiFi channels are added to the resource pool of a BS at rate $q_1^{[i]}$ and released at rate $q_3^{[i]}$. LTE PRBs are instead acquired at rate $q_2^{[i]}$ and released at rate $q_4^{[i]}$. In Section V such rates will be used to evaluate three different resource allocation policies.

More details on the rates $r_3^{[i]}, \dots, r_{12}^{[i]}$, and $q_1^{[i]}, \dots, q_4^{[i]}$, are presented in the Appendix.

IV. MODEL SOLUTION

We solve the above model as follows. The agent presented in Fig. 2 is used to determine, for each coverage area L_i , the number of states. In particular, each area L_i , with $n_i = |\text{neigh}(i)|$ neighbor, will be described by $2n_i + 5$ states. Each state s of a coverage area L_i is described by a variable $n_s^{[i]}$ counting the number of agents in the considered area in that state. We perform a *mean field* approximation of the number of agents in a given state. In other words, instead of taking into account the evolution of the distribution of states in which each agent could be, we only consider the mean number of agents in each state. Note that, even if at any given time instant the number of agents is discrete, its mean number is continuous, thus we can consider variables $n_s^{[i]}$ as positive continuous quantities. The temporal evolution of the system can then be evaluated using two matrices and a vector for each area L_i , whose elements are functions of the variables $n_s^{[i]}$. The *transition matrix* $\mathbf{C}^{[i]} = |c_{su}^{[i]}|$ contains the transition rates $c_{su}^{[i]}$ from state s to state u , which can be obtained by using the rates defined in Section III-A. The *death matrix*, $\mathbf{D}^{[i]} = \text{diag}(d_{ss}^{[i]})$, is a diagonal matrix whose elements $d_{ss}^{[i]}$ represent agents that leave the system in state s (in our case, they correspond to successful request transmissions). Finally, the *birth vector* $\mathbf{b}^{[i]} = |b_s^{[i]}|$ represents the arrival of new agents in state s . As for the transition matrix $\mathbf{C}^{[i]}$, also the elements of $\mathbf{D}^{[i]}$ and $\mathbf{b}^{[i]}$ can be computed using the rates presented in Section III-A.

The number of users in a given state is collected in a row vector $\mathbf{n}^{[i]} = |n_s^{[i]}|$, and the evolution of the system is obtained by solving the following coupled ordinary differential equations¹ for each coverage area L_i ($i = 0, \dots, n - 1$):

$$\frac{d\mathbf{n}^{[i]}}{dt} = \mathbf{n}^{[i]} \left(\mathbf{C}^{[i]} - \mathbf{D}^{[i]} \right) + \mathbf{b}^{[i]}. \quad (3)$$

Equations (3) can be solved using suitable numerical algorithms, such as the *Runge-Kutta with adaptive step-size control* discretization method [19] used in this work.

For example, let us focus on a user in area L_2 in Fig. 1, and let us assume that the states are ordered as follows: *WiFi local*, *WiFi rem₁*, *WiFi rem₃*, *LTE local*, *LTE rem₁*, *LTE rem₃*, *Loss*. BS₂ has two neighbors, that is: $\text{neigh}(2) = \{1, 3\}$. Then, the resulting matrices and vectors

¹Even if all vectors and matrices depend on time, we have omitted the explicit dependency in order to simplify the notation.

are given by:

$$\mathbf{C}^{[2]} = \begin{pmatrix} -() & r_{6:1}^{[2]} & r_{6:3}^{[2]} & 0 & 0 & 0 & 0 \\ r_{10:1}^{[2]} & -() & 0 & r_{7:1}^{[2]} & r_{8:1,1}^{[2]} & r_{8:1,3}^{[2]} & r_{9:1}^{[2]} \\ r_{10:3}^{[2]} & 0 & -() & r_{7:3}^{[2]} & r_{8:3,1}^{[2]} & r_{8:3,3}^{[2]} & r_{9:3}^{[2]} \\ r_{11}^{[2]} & 0 & 0 & -() & 0 & 0 & 0 \\ r_{12:1}^{[2]} & 0 & 0 & -() & 0 & 0 & 0 \\ r_{12:3}^{[2]} & 0 & 0 & -() & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{D}^{[2]} = \text{diag} \begin{pmatrix} r_2^{[2]} \\ r_{3:1}^{[2]} \\ r_{3:3}^{[2]} \\ r_4^{[2]} \\ r_{5:1}^{[2]} \\ r_{5:3}^{[2]} \\ 0 \end{pmatrix} \quad \mathbf{b}^{[2]} = \begin{pmatrix} r_1^{[2]} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}^T$$

Since matrix $\mathbf{C}^{[i]}$ must be an infinitesimal generator (that is, all its rows must sum up to zero), we have used the notation “ $-()$ ” to identify the sum of the other elements in the row, changed in sign. Note that the last row is composed by all 0s, since the *Loss* state is absorbing.

A. Computation of performance metrics

From the evolution of the number of agents $\mathbf{n}^{[i]}$ in a given state for each area L_i , we can derive several indices that can be used to assess the system performance. The transition rate $r_{9:j}^{[i]}$ expresses the frequencies at which requests are lost due to limited resources. Thus, we can compute the loss rate Λ_i at cell L_i as:

$$\Lambda_i = \sum_{j \in \text{neigh}(i)} r_{9:j}^{[i]}.$$

Note also that $n_{Loss}^{[i]}$ represents the total number of user requests lost in area L_i in the considered time interval T , that is:

$$n_{Loss}^{[i]} = \int_0^T \Lambda_i dt$$

which gives us insights on the number of transmission failures that the users have experienced. The number of requests enqueued for transmission at BS_i for a given technology, $Q_{WiFi}^{[i]}$ and $Q_{LTE}^{[i]}$, can be obtained directly from the elements of $\mathbf{n}^{[i]}$:

$$Q_{WiFi}^{[i]} = n_{WiFi\ local}^{[i]} + \sum_{j \in \text{neigh}(i)} n_{WiFi\ rem_i}^{[j]}$$

$$Q_{LTE}^{[i]} = n_{LTE\ local}^{[i]} + \sum_{j \in \text{neigh}(i)} n_{LTE\ rem_i}^{[j]}.$$

Note that, since BS_i can serve also some of the users at the neighboring locations L_j , the total number of enqueued requests at BS_i must consider such requests as well (terms $n_{WiFi\ rem_i}^{[j]}$ and $n_{LTE\ rem_i}^{[j]}$). Then, the total number of requests waiting to be served in the system can simply be computed as $Q_{WiFi} = \sum_i Q_{WiFi}^{[i]}$ and $Q_{LTE} = \sum_i Q_{LTE}^{[i]}$.

Throughput corresponding to the BS_i interfaces, denoted by $X_{WiFi}^{[i]}$ and $X_{LTE}^{[i]}$, can be computed in a similar way. Specifically, taking into account the speed at which each request is served as well as the number of requests being served, we can write

$$X_{WiFi}^{[i]} = n_{WiFi\ local}^{[i]} \sigma_{WiFi}^{[i]} + \sum_{j \in \text{neigh}(i)} n_{WiFi\ rem_i}^{[j]} \sigma_{WiFi}^{[j]}$$

$$X_{LTE}^{[i]} = n_{LTE\ local}^{[i]} \sigma_{LTE}^{[i]} + \sum_{j \in \text{neigh}(i)} n_{LTE\ rem_i}^{[j]} \sigma_{LTE}^{[j]}$$

where $\sigma_{WiFi}^{[i]}$ and $\sigma_{LTE}^{[i]}$ were given in, respectively, (1) and (2).

V. PERFORMANCE EVALUATION

We now study the network presented in Fig. 1 in presence of time varying high traffic loads, under three different radio resource allocation policies. The network we consider is composed by eight BSs, and each of them has two neighboring BSs. The load varies in time and in space, creating bursts in different locations of the network at different times (the so-called flash-crowd events). The resource allocation policies under study are as follows. The *Uniform* policy assigns to every BS a constant and equal number of WiFi and LTE resources. The *Static* one allocates to all BSs a constant but non homogeneous number of radio resources; the number depends on the traffic load distribution experienced in the past, i.e., BSs that have experienced a high traffic load in the past get more resources. The *Dynamic* policy allocates resources according to the actual load experienced by a BS over time. Given that the total number of resources is fixed, when a BS gets more resources under the Static and Dynamic policies, some other BSs will get fewer. The comparison among the three policies is based on the computation of the following performance indexes, presented in Section IV-A:

- number of requests lost by the network;
- WiFi average request queue;
- LTE average request queue;
- WiFi throughput for each BS;
- LTE throughput for each BS.

A. Setting and results

The network load is characterized by three request bursts of 12 accesses per second by users within coverage areas L_2 , L_4 , and L_7 . Specifically, the request burst affects BS_2 from time instant 0 s to 50 s, BS_4 from 70 s to 120 s, and BS_7 from 140 s to 190 s. The peak load is typically higher than the maximum capacity that a BS can handle: this creates queues and, possibly, packet losses. Under regular traffic load conditions, each BS receives 0.05 access requests per second. The network load for BSs 0, 2, 4, and 7 is reported in Fig. 5 (dotted lines).

The Uniform policy assigns to each BS 4 WiFi channels and 25 LTE PRBs (per subframe). The Static allocates 3 more radio WiFi channels (total 7) to BS_2 and BS_4 , by taking these channels from their neighboring BSs. In a similar way,

it allocates 15 more LTE PRBs (for a total of 40 PRB per subframe) to BS₂ and BS₄. The third burst experienced by BS₇ is not foreseen by the Static policy, hence no resource increase is performed at that BS. The Dynamic policy leads to an increase in the radio resources at all overloaded BSs, including BS₇, but only when a request burst occurs.

Fig. 3 depicts the time evolution of the total number of lost requests under each policy. As expected, the Uniform policy has the largest number of losses due to the overloading of BS₂, BS₄ and BS₇. The Dynamic outperforms all other policies, as it leads to the least total number of dropped requests. The fact that, from 120 s to 170 s, the Static policy implies fewer losses than the Dynamic is a consequence of the delay in dynamically allocating resources to BSs that become overloaded. However, the unexpected burst at BS₇ clearly causes a higher number of losses under the Static policy (170 – 200 s time interval).

In Fig. 4, we present the WiFi and LTE queue average occupancy. The left plot shows that, under the Uniform policy, less resources are available resulting in a shorter queue, i.e., many access requests are dropped. When the third burst arrives, the Dynamic policy is the only one that provides enough WiFi resources for handling the sudden surge in the traffic load. Hence, a longer WiFi queue is observed. Again, we remark that a longer queue does not mean a longer response time, rather that fewer requests are dropped. The right plot indicates that during the third burst the LTE queue is shorter under the Dynamic policy, since more WiFi resources are available and fewer switches to LTE are necessary (note that curves referring to Static and Uniform overlap).

Finally, Fig. 5 depicts the incoming access requests and the WiFi/LTE throughput for a BS with regular traffic load, namely, BS₀, and for the overloaded BSs (namely, 2, 4 and 7) when the Dynamic policy is implemented. As it can be seen, when the request queue at the WiFi interface becomes too long (both at the local and at the neighboring BSs), LTE starts being used. After the burst is over, both the LTE and the WiFi technologies are still fully used for a bit longer in order to let all pending requests be served. BS₀, which is not affected by any traffic burst, experiences an increased load because it takes part of the neighboring BSs load on itself. Specifically, it serves some requests generated by users within L_7 (for $t > 150$ s) and within L_2 (for $t < 50$ s).

VI. CONCLUSIONS AND FUTURE WORK

We presented an analytical framework based on the Markovian agent formalism, which models multi-RAT cellular networks. We envisioned a system where base stations can provide Internet connectivity through the LTE as well as the WiFi technology, and we highlighted how our framework can model the different dynamics of the system reflecting the user traffic. By solving the model through a mean-field based methodology, we also showed that the framework can well capture the system behavior in flash-crowd scenarios. Future work will focus on the model validation through simulation, as well as on the study of different user connectivity strategies.

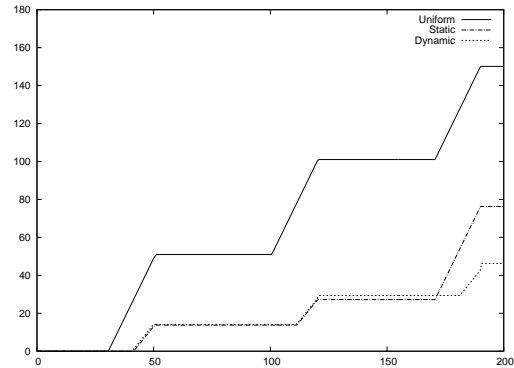


Fig. 3. Policy comparison: time evolution of the number of lost access requests.

REFERENCES

- [1] Architecture for mobile data offload over Wi-Fi access networks. http://www.cisco.com/en/US/solutions/collateral/ns341/ns524/ns673/white_paper/c11-701018.html. [Online; accessed February-2013].
- [2] Mobile data offloading for 3g and lte networks. <http://www.alvarion.it/applications/mobile-data-offloading>. [Online; accessed February 2013].
- [3] Economy + internet trends: Web 2.0 summit. http://www.morganstanley.com/institutional/techresearch/pdfs/MS_Economy_Internet_Trends_102009_FINAL.pdf, 2009.
- [4] Cisco Visual Networking Index: Global mobile data traffic forecast update, 2011–2016, Cisco White Paper, feb. 2012.
- [5] Aruna Balasubramanian, Ratul Mahajan, and Arun Venkataramani. Augmenting mobile 3g using wifi. In *Proceedings of the 8th international conference on Mobile systems, applications, and services, MobiSys '10*, pages 209–222. ACM, 2010.
- [6] M. Benaim and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, 2008.
- [7] A. Bobbio, M. Griboaldo, and 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.
- [8] S. Busanelli, M. Martalò, G. Ferrari, G. Spigoni, and N. Iotti. Vertical handover between wifi and umts networks: Experimental performance analysis. *International Journal of Energy, Information and Communications*, 2011.
- [9] F. Cordero, D. Manini, and M. Griboaldo. 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)*, pages 193–211. IEEE Computer Society Press, 2009.
- [10] S. Dimatteo, Pan Hui, Bo Han, and V.O.K. Li. Cellular traffic offloading through wifi networks. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 192–201, oct. 2011.
- [11] H.M. ElBadawy. Optimal rat selection algorithm through common radio resource management in heterogeneous wireless networks. In *Radio Science Conference (NRSC), 2011 28th National*, pages 1–9, 2011.
- [12] M. Griboaldo, D. Cerotti, and 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.
- [13] Marco Griboaldo, Daniele Manini, and Carla-Fabiana Chiasserini. Studying mobile internet technologies with agent based mean-field models. In *ASMTA*, pages 112–126, 2013.
- [14] David Hadaller, Srinivasan Keshav, Tim Brecht, and Shubham Agarwal. Vehicular opportunistic communication under the microscope. In *Proceedings of the 5th international conference on Mobile systems, applications and services, MobiSys '07*, pages 206–219. ACM, 2007.
- [15] Bo Han, Pan Hui, V.S.A. Kumar, M.V. Marathe, Jianhua Shao, and A. Srinivasan. Mobile data offloading through opportunistic communications and social participation. *Mobile Computing, IEEE Transactions on*, 11(5):821–834, may 2012.

- [16] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11b. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 836 – 843 vol.2, march-3 april 2003.
- [17] Harri Holma and Antti Toskala. *LTE for UMTS Evolution to LTE-Advanced*. John Wiley and Sons Ltd, 2011.
- [18] Bret Hull, Vladimir Bychkovsky, Yang Zhang, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Hari Balakrishnan, and Samuel Madden. Cartel: a distributed mobile sensor computing system. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys '06, pages 125–138. ACM, 2006.
- [19] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007.
- [20] A.A. Sabbagh, R. Braun, and M. Abolhasa. A comprehensive survey on rat selection algorithms for heterogeneous networks. *World Academy of Science, Engineering and Technology*, 49:141–145, 2011.

APPENDIX

In this appendix, we show the transition rates of the technologies and cell switching to perform either off-loading or upgrading. All rates uses the following support functions to determine the neighbor BS when a cell switching is involved. In particular, we call $I_C(cond)$ the indicator function, that returns 1 if the condition is *true*, and 0 if the condition is false. We also define $I_M(\dots)$ as follows:

$$I_M(j : v_1, \dots, v_j, \dots) = \begin{cases} 0 & \text{if } v_j > \min(v_i) \\ 1 & \text{if } v_j = \min(v_i) \text{ and} \\ & v_k > v_j \forall k \neq j \\ \frac{1}{n} & \text{if } v_j = \min(v_i) \text{ and} \\ & \text{there are } n \text{ } v_k : v_k = \min(v_i) \end{cases} \quad (4)$$

In other words, function $I_M(j : \dots)$ returns 1 if its argument v_j is minimum among the other arguments, 0 otherwise. However if there other arguments equal to v_j at the minimum (n in total), the function returns $1/n$ to equally share the decision among the available candidate to simulate a random selection policy.

WiFi switching to neighboring BSs.

$$r_{6:j}^{[i]} = \mu_{BS-Sw} \cdot I_C(\sigma_{WiFi}^{[i]} < \mu_{min}) \cdot I_M(j : \sigma_{WiFi}^{[neigh(i)]})$$

where the notation $I_M(j : \sigma_{WiFi}^{[neigh(i)]})$ means that function $I_M(\dots)$ defined above is applied to all the BS_k with $k \in neigh(i)$.

Switching from neighboring WiFi to LTE.

$$r_{7:j}^{[i]} = \mu_{\tau-Sw} \cdot I_C(\sigma_{WiFi}^{[j]} < \mu_{min}) \cdot I_C(Q_{LTE}^{[i]} < \gamma_{min} \cdot n_{LTE}^{[i]} resources)$$

$$r_{8:j,j'}^{[i]} = \mu_{\tau-Sw} \cdot I_C(\sigma_{WiFi}^{[j]} < \mu_{min}) \cdot I_C(Q_{LTE}^{[i]} > \gamma_{min} \cdot n_{LTE}^{[i]} resources) \cdot I_C(Q_{LTE}^{[j']} < \gamma_{min} \cdot n_{LTE}^{[j']} resources) \cdot I_M(j' : \sigma_{WiFi}^{[neigh(i)]})$$

Loss of requests.

$$r_{9:j}^{[i]} = \mu_{\tau-Sw} \cdot I_C(\sigma_{WiFi}^{[j]} < \mu_{min}) \cdot I_C(Q_{LTE}^{[i]} > \gamma_{min} \cdot n_{LTE}^{[i]} resources) \cdot \prod_{j' \in neigh(i)} I_C(Q_{LTE}^{[j']} > \gamma_{min} \cdot n_{LTE}^{[j']} resources)$$

Upgrade.

$$r_{10;j}^{[i]} = \mu_{BS-Sw} \cdot I_C(\sigma_{WiFi}^{[i]} > \mu_{max}) \\ r_{11}^{[i]} = r_{12;j}^{[i]} = \mu_{\tau-Sw} \cdot I_C(\sigma_{WiFi}^{[i]} > \mu_{max})$$

Note that upgrade occurs at different speeds, depending on whether it involves technology switch or cell switch.

Radio resource allocation.

$$q_1^{[i]} = \eta \cdot \rho_{WiFi}^{[i]}, \quad q_2^{[i]} = \eta \cdot \rho_{LTE}^{[i]}, \quad q_3^{[i]} = q_4^{[i]} = \eta$$

where $\rho_{WiFi}^{[i]}$ and $\rho_{LTE}^{[i]}$ are the target resource level for the considered policy at BS_i . Note that due to Equation 3, the system will converge to the target number of resource $\rho_{WiFi}^{[i]}$ and $\rho_{LTE}^{[i]}$ at a speed determined by parameter η .

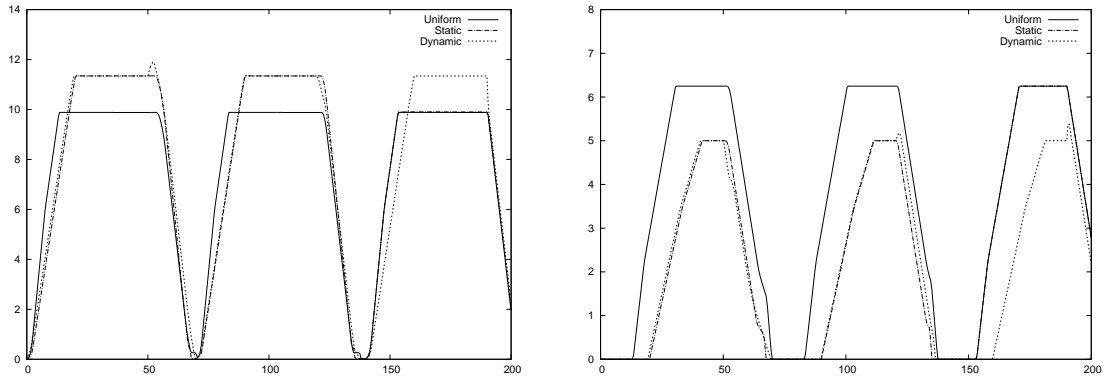


Fig. 4. Time evolution of WiFi queue (left) and LTE queue (right).

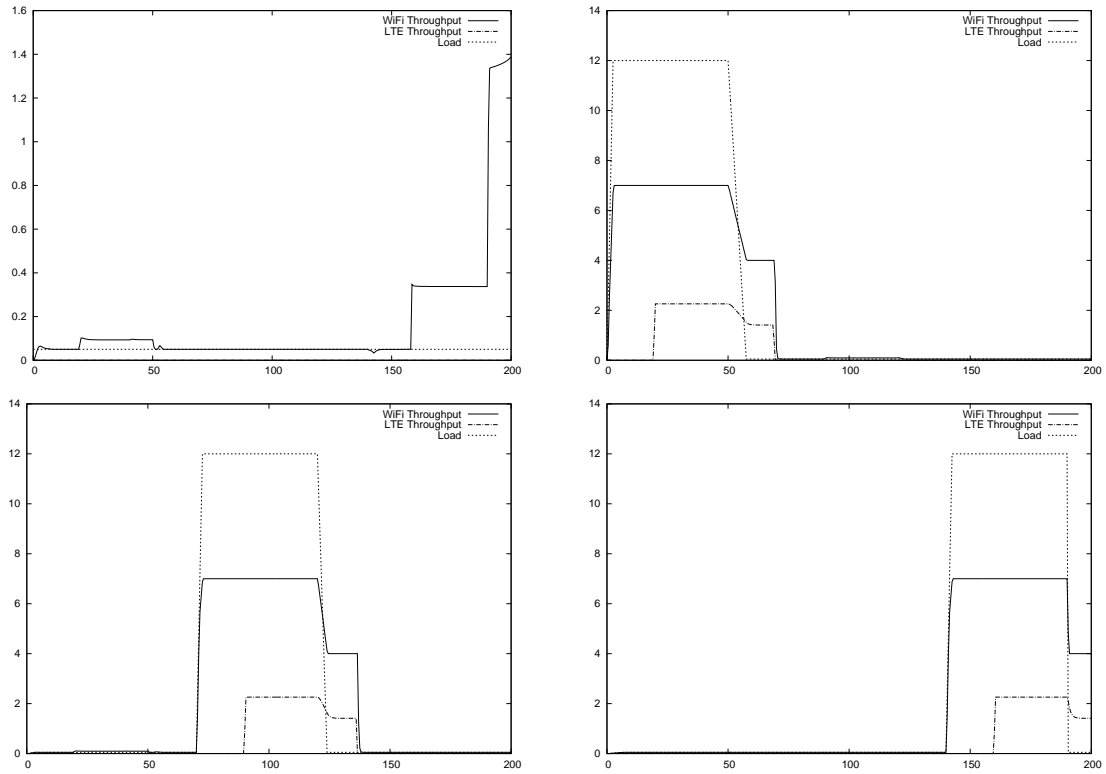


Fig. 5. Dynamic policy: time evolution of traffic load, WiFi throughput and LTE throughput for cells 0 (top left), 2 (top right), 4 (bottom left) and 7 (bottom right).