Performance Analysis of Data Traffic in Small Cells Networks with User Mobility

Philippe Olivier, Florian Simatos and Alain Simonian

Abstract We analyze the impact of inter-cell mobility on data traffic performance in dense networks with small cells. To this end, a multi-class queuing system with impatience is proposed as a generic model that captures mobility through the sojourn time of users in the cell. We provide mathematical proofs for the stability and the regularity of this multi-class queuing system. We then show how the performance of a homogeneous network is amenable to the application of the generic model to a single representative cell. This model is applied to derive the throughput of both mobile and static users, along with the handover probability. Numerical evaluation and simulation results are provided to assess the accuracy of the approach; we show, in particular, that both classes of users benefit from a throughput gain induced by the opportunistic displacement of mobile users among cells.

Key words: Cellular Networks; Mobility; Traffic; Performance Evaluation; Markov Processes; Queue Stability

1 Introduction

To address the permanent increase of mobile traffic, the capacity of networks can be upgraded by a massive deployment of small cells. This solution is notably envisaged by network operators for the LTE-A heterogeneous networks [9] or Ultra Dense Networks scenarios for future 5G networks [15]. In dense networks, however, the amount of handover generated by users mobility will increase with a notable impact on signaling overhead, and possibly on the throughput of data transfers.

Philippe Olivier, Alain Simonian
Orange Labs, 44 avenue de la République, 92320 Châtillon, France, e-mail: phil.olivier@orange.com, alain.simonian@orange.com

Florian Simatos
ISAE Supaero, 10 avenue Edouard Belin, 31000 Toulouse, France, e-mail: florian.simatos@isae.fr

In this context, the present paper aims at evaluating the impact of *inter-cell mobility* on the performance of data traffic in dense networks. Specifically, considering small cells enables us to neglect the possible spatial variations of intra-cell capacities and thus to focus on the impact of inter-cell mobility itself. Furthermore, we decouple the performance evaluation problem from the modeling of the user displacement in the plane, the latter topic being out of scope of the present paper (see [13, 17] for current displacement models). Mobility is here supposed to be captured through the distribution of the users residual sojourn time in a cell, that is, the time a mobile user is physically present in the cell once its transmission has started.

Given this distribution, we construct a flow-level queuing model that allows us to derive the essential performance metrics in each cell, namely the mean throughput and the handover probability. The generic tool of our model is a multi-class Processor Sharing (PS) queue with “impatient” customers; the impatience here accounts for the mobility of customers from cell to cell. This generic model can then be applied to each individual cell to solve the set of flow equations, which characterize the handover rates between cells, and compute the performance indicators.

To our knowledge, the PS queue with impatience has been mainly addressed in terms of asymptotic regimes for the reneging probability for one customer class [8] or for several classes in overload [10]. The analysis of the stable multi-class PS queue with distinct impatience rates, however, has not received so far a significant contribution. For this multi-class queuing system, we here provide proofs for the stability condition and for regularity properties of the empty-system probability.

Throughput gains induced by mobility in cellular networks have been generally related so far to the spatial variations of capacity inside the cells, which permits an opportunistic use of favorable transmission conditions by mobile users [1, 4, 5, 7, 11]. These papers base their evaluation on flow-level traffic modeling and address mobility through a spatial Markov process where users jump between distinct capacity zones in the cells. Due to the complexity of the latter approach, performance indicators can be derived through suitable bounds or approximations only. In the present work, by decoupling the queuing and mobility models, we alternatively formulate the problem in terms of an equilibrium regime for the handover flows, the existence of which is assessed in the case of a homogeneous network.

The paper is organized as follows. A generic one-cell Markovian model is first constructed in Section 2 and the stability and regularity properties are stated and proved; Section 3 presents our approach to model networks with mobility; Section 4 presents numerical results, including simulation, and their discussion; finally, Section 5 draws conclusions and summarizes our main achievements.

2 Generic Queueing Model

As a first step, we consider a single cell model which is used as the generic tool to further analyze the impact of inter-cell mobility in a network.

2.1 A PS Queue with Impatience

The considered cell is supposed to be "small", i.e., of limited range so that its transmission capacity C can be assumed spatially constant; this capacity is viewed as an input parameter accounting for radio and interference conditions in the considered cellular network. We suppose that capacity C is equally shared among all active users present in its service area, as implemented by means of a Round-Robin scheduling. Following this fair sharing policy, the system occupancy at the flow level can then be modeled by a Processor-Sharing (PS) queue [6].

We consider K classes of users which generate requests for transmission according to Poisson processes with respective arrival rate λ_k, $k = 1,...,K$. Class-k users have i.i.d. transmission requests of data volume Σ_k with mean σ_k, hence a service rate $\mu_k = C/\sigma_k$. Since customers may actually leave the cell during their communications, we call T_k the *remaining* sojourn time of a mobile user, i.e., the time duration he physically stays in the cell after the transmission has started. We finally denote by $\theta_k = 1/E(T_k)$ the mean cell departure rate of class-k users, called class-k *mobility rate*; any class k where $\theta_k = 0$ will be called *static*. The cell occupancy can then be described by the K-dimensional process $\mathbf{N}(t) = (N_1(t),...,N_K(t))$, $t \geq 0$, where $N_k(t)$ denotes the number of ongoing class-k data transfers at time t. This process evolves as the occupancy of a PS queue with impatience, the "impatient" customers here corresponding to mobile users that may leave the system before their service completion within the given cell.

We assume that Σ_k and T_k are exponentially distributed with parameters $1/\sigma_k$ and θ_k, respectively. The process $(\mathbf{N}(t))_{t\geq 0}$ is then Markovian in the state space $\mathbb{N}^K$; from state $\mathbf{n} = (n_1,...,n_K)$ and for $\mathbf{e}_k = (0,...,1,...0)$ with 1 at the k-th component, it can reach state $\mathbf{n}+\mathbf{e}_k$ with transition rate λ_k, or state $\mathbf{n}-\mathbf{e}_k$ with transition rate $n_k\mu_k/L(\mathbf{n}) + n_k\theta_k$, denoting by $L(\mathbf{n}) = \sum_{1\leq j\leq K} n_j$ the total number of active users. Let $\rho_k = \lambda_k/\mu_k$ be the offered load of class k and S (resp. M) denote the set of static (resp. mobile) classes.

In stationary regime , the equilibrium equations of process $(\mathbf{N}(t))_{t\geq 0}$ read

$$\sum_{k=1}^{K}\left[\lambda_k + n_k\left(\frac{\mu_k}{L(\mathbf{n})}+\theta_k\right)\right]\mathbb{P}(\mathbf{N}=\mathbf{n}) = \sum_{k=1}^{K}\lambda_k\,\mathbb{P}(\mathbf{N}=\mathbf{n}-\mathbf{e}_k) + \sum_{k=1}^{K}(n_k+1)\left(\frac{\mu_k}{L(\mathbf{n})+1}+\theta_k\right)\mathbb{P}(\mathbf{N}=\mathbf{n}+\mathbf{e}_k) \tag{1}$$

with $\sum_{n\in\mathbb{N}^K}\mathbb{P}(\mathbf{N}=\mathbf{n}) = 1$. Process $(\mathbf{N}(t))_{t\geq 0}$ is not reversible unless all classes are static; its stationary distribution is thus not amenable to a simple closed form. Nevertheless, a general conservation law between the average arrival and departure rates can be stated as follows: for given k, multiplying each equation of (1) by n_k and then summing over all state vectors $\mathbf{n} \in \mathbb{N}^K$ provides

$$\lambda_k = \mu_k\,\mathbb{E}\left(\frac{N_k\,\mathbf{1}_{N_k>0}}{L(\mathbf{N})}\right) + \theta_k\,\mathbb{E}(N_k), \quad 1 \leq k \leq K. \tag{2}$$

Proposition 2.1 ***The Markov process*** $(\mathbf{N}(t))_{t\geq 0}$ ***has a stationary regime if and only if***

$$\rho_S = \sum_{k\in S} \rho_k < 1. \tag{3}$$

Proof. First assume that process $(\mathbf{N}(t))_{t\geq 0}$ has a stationary distribution; applying conservation law (2) to each static class k with $\theta_k = 0$, then summing over all $k \in S$, gives $\rho_S = \sum_{k\in S} \rho_k = \sum_{k\in S} \mathbb{E}(N_k \mathbf{1}_{N_k>0}/L(\mathbf{N})) < 1$, so that condition (3) is necessary.

Conversely, assume that (3) holds. For any test function $f : \mathbb{N}^K \to \mathbb{R}^+$, the infinitesimal generator $\mathscr{Q}$ of the Markov process $(\mathbf{N}(t))_{t\geq 0}$ is given by

$$\mathscr{Q}f(\mathbf{n}) = \sum_{1\leq k\leq K} \lambda_k \left[f(\mathbf{n}+\mathbf{e}_k) - f(\mathbf{n})\right] + \\ \sum_{1\leq k\leq K} \left(\frac{\mu_k n_k}{L(\mathbf{n})} + \theta_k n_k\right) \mathbf{1}_{n_k>0} \left[f(\mathbf{n}-\mathbf{e}_k) - f(\mathbf{n})\right], \quad \mathbf{n} \in \mathbb{N}^K.$$

By ([16], Proposition 8.14), the process $(\mathbf{N}(t))_{t\geq 0}$ is ergodic if there exists a so-called Lyapunov function $\Lambda : \mathbb{N}^K \to \mathbb{R}^+$ and positive constants η, δ such that

(a) the set $\{\mathbf{n} \in \mathbb{N}^K, \Lambda(\mathbf{n}) \leq \eta\}$ is finite,
(b) random variables $\sup_{0\leq t\leq 1} \Lambda(\mathbf{N}(t))$ and $\int_{[0,1]} |\mathscr{Q}\Lambda(\mathbf{N}(t))| \mathrm{d}t$ are integrable,
(c) $\mathscr{Q}\Lambda(\mathbf{n}) \leq -\delta$ as soon as $\Lambda(\mathbf{n}) > \eta$.

Consider the function $\Lambda : \mathbf{n} \in \mathbb{N}^K \mapsto \Lambda(\mathbf{n})$ defined by $\Lambda(\mathbf{n}) = s^2 + m^2$ with $s = \sum_{i\in S} n_i/\mu_i$, $m = \sum_{j\in M} n_j/\mu_j$. We successively verify conditions (a), (b) and (c):

• (a) is clearly fulfilled by Λ and any finite η;

• if $A_k(t)$ is the number of class-k user arrivals within interval $[0,t]$, we readily have $N_k(t) \leq A_k(t) \leq A_k(1)$ for $0 \leq t \leq 1$, where variable $A_k(1)$ has finite first and second moments. The latter inequalities thus ensure the validity of (b) for Λ;

• denoting by $\rho_M = \sum_{j\in M} \rho_j$ the mobile load, the above definition of $\mathscr{Q}$ yields

$$\mathscr{Q}\Lambda(\mathbf{n}) = \sum_{1\leq k\leq K} \frac{\rho_k}{\mu_k} + \frac{1}{L(\mathbf{n})} \sum_{1\leq k\leq K} \frac{n_k}{\mu_k} + 2(\rho_S - 1)s + \\ \frac{2(s-m)}{L(\mathbf{n})} \sum_{j\in M} n_j + 2m \left(\rho_M - \sum_{j\in M} n_j \frac{\theta_j}{\mu_j}\right) + \sum_{j\in M} n_j \frac{\theta_j}{\mu_j^2}$$

for $\mathbf{n} \neq \mathbf{0}$. Setting $\mu_* = \min_{1\leq k\leq K} \mu_k$, $\mu^{**} = \max_{1\leq k\leq K} \mu_k$ and $A = \min_{j\in M} \theta_j/\mu_j$, $B = \max_{j\in M} \theta_j/\mu_j^2$, we then derive the upper bound

$$\mathscr{Q}\Lambda(\mathbf{n}) \leq \sum_{1\leq k\leq K} \frac{\rho_k}{\mu_k} + \frac{1}{\mu_*} + 2(\rho_S - 1)s + m\left[2\rho_M + 2\frac{\mu^{**}}{\mu_*} + B\mu^{**}\right] - 2A\mu_* m^2. \tag{4}$$

As $\rho_S < 1$ by condition (3), we deduce from (4) that $\mathscr{Q}\Lambda(\mathbf{n})$ is asymptotically smaller than $-2A\mu_{min}m^2$ when m tends to infinity. Thus, for any given $\delta > 0$, there exists a constant $m_0 > 0$ such that $\mathscr{Q}\Lambda(\mathbf{n}) < -\delta$ as soon as $m > m_0$. Now,

- if $s \leq m$, it is sufficient that $s+m > 2m_0$ to ensure that $m > m_0$;
- if $s > m$ and $m \leq m_0$, all terms depending on m in (4) are bounded and $\mathscr{Q}\Lambda(\mathbf{n})$ is then asymptotically smaller than $2(\rho_S - 1)s$ when s tends to infinity.

There exists thus a constant $s_0 > 0$ such that $\mathscr{Q}\Lambda(\mathbf{n}) < -\delta$ as soon as $s > s_0$. Fixing the constant $\eta = (\max(2s_0, 2m_0))^2$ and using $(s+m)^2 \geq s^2 + m^2$, we conclude that $\mathscr{Q}\Lambda(\mathbf{n}) < -\delta$ when $\Lambda(\mathbf{n}) > \eta$, thus fulfilling requirement (c).

Conditions (a), (b) and (c) being verified, Λ is therefore a Lyapunov function for process $(\mathbf{N}(t))_{t \geq 0}$ and condition (3) is thus also sufficient. □

Note that condition (3) does not depend on the traffic intensity of mobile users, since the latter leave the cell after a finite time and thus cannot cause overload. Now, given (3), we define two performance indicators per user-class, the *average throughput* and the *handover probability*. Considering data (elastic) traffic, the user-perceived QoS can be measured by the average throughput defined as the ratio of the mean volume of transferred data to the mean transfer time [6]. We also define the *handover probability* for class-k users as the proportion of users that exit the cell before the completion of their transmission, i.e., the ratio of the mean handover rate λ_k^{Out} to the mean flow arrival rate λ_k. The latter definitions read

$$\Gamma_k \triangleq \frac{\mathbb{E}(X_k)}{\mathbb{E}(\Delta_k)}, \quad H_k \triangleq \frac{\lambda_k^{Out}}{\lambda_k}, \qquad 1 \leq k \leq K, \tag{5}$$

where X_k denotes the part of the total data volume Σ_k which is actually transferred by a class-k user during its transmission time Δ_k $(\leq T_k)$ in the cell. The following proposition is easily derived, which proof has been given in [14].

Proposition 2.2 ***The throughput Γ_k and the handover probability H_k are given by***

$$\Gamma_k = C\left(\frac{\rho_k}{\mathbb{E}(N_k)} - \frac{\theta_k}{\mu_k}\right), \quad H_k = \frac{\mathbb{E}(N_k)\,\theta_k}{\lambda_k}, \quad 1 \leq k \leq K \tag{6}$$

which depend on the mean number of class-k users only. They satisfy the remarkable identity

$$H_k = \frac{\theta_k\,\sigma_k}{\Gamma_k + \theta_k\,\sigma_k}, \quad 1 \leq k \leq K.$$

2.2 Regularity Properties of the Empty-system Probability

Monotonicity and continuity of the empty-system probability as a function of any arrival rate λ_k, will prove essential in Section 3 to solve the equilibrium equations of handover flows in a network. We claim that such regularity properties require a specific proof in the present queuing system with infinite state space and no closed form solution for the stationary distribution.

Denote by $\mathscr{A}_K(\lambda_K)$ the PS queuing system with impatience and define

$$Q(\lambda_K) = \mathbb{P}(\mathbf{N} = \mathbf{0}), \quad \lambda_K \geq 0, \tag{7}$$

the empty-queue probability as a function of the rate λ_K of class-K users, all other parameters kept constant (distinguishing here class K, be it a static class or not).

Proposition 2.3 ***Function $Q(.)$ is strictly decreasing over its definition interval.***

Proof. The definition domain of function Q corresponds to those values of λ_K such that $\rho_S < 1$, according to (3). To prove the proposition, we proceed in four steps.

A) We compare the empty queue probabilities $Q(\lambda_K)$ and $Q(\lambda_K')$ of systems $\mathscr{A}_K(\lambda_K)$ and $\mathscr{A}_K(\lambda_K')$ with $\lambda_K' = \lambda_K + \Delta\lambda_K$, $\Delta\lambda_K > 0$. To do this, it proves convenient to introduce a supplementary user class by defining a new system $\overline{\mathscr{A}}_{K+1}$ with $K+1$ classes, where the first K classes are identical to that in $\mathscr{A}_K(\lambda_K)$ and where the $(K+1)$th class has input rate $\lambda_{K+1} = \Delta\lambda_K$, service rate $\mu_{K+1} = \mu_K$ and impatience rate $\theta_{K+1} = \theta_K$. The occupancy of system $\overline{\mathscr{A}}_{K+1}$ is now defined by the $(K+1)$-dimensional vector $\overline{\mathbf{N}}(t) = (\overline{N}_1(t), ..., \overline{N}_K(t), \overline{N}_{K+1}(t))$, $t \geq 0$.

It proves that the K-dimensional Markov occupancy process, deduced from system $\overline{\mathscr{A}}_{K+1}$ by gathering populations of classes K and $K+1$, has the same transition rates, and thus the same stationary distribution, as the process $(\mathbf{N}'(t))_{t\geq 0}$ of system $\mathscr{A}_K(\lambda_K')$. This result holds essentially because of the Poisson nature of all the arrival processes and of the "PS + Impatience" form of the departure processes, for which the departure rates do not depend on the fact that some classes are gathered or not.

System $\overline{\mathscr{A}}_{K+1}$ (with occupancy $\overline{\mathbf{N}}$) can therefore be considered in place of system $\mathscr{A}_K(\lambda_K')$ for the evaluation of the stationary empty-system probability, and we have

$$Q(\lambda_K') = \mathbb{P}(\overline{\mathbf{N}} = \mathbf{0}). \tag{8}$$

B) We now make use of a sample path argument to state that, at any time, each population $N_k(t)$, $1 \leq k \leq K$, is no greater than the corresponding population $\overline{N}_k(t)$, assuming that both systems $\mathscr{A}_K(\lambda_K)$ and $\overline{\mathscr{A}}_{K+1}$ are empty at $t = 0$ (for convenience, we set $N_{K+1}(t) = 0$ for all $t \geq 0$).

This may be thoroughly proved by induction on the sequence of all consecutive events (arrivals or departures) occurring in either system $\mathscr{A}_K(\lambda_K)$ or $\overline{\mathscr{A}}_{K+1}$. The result essentially holds because 1) there are supplementary arrivals in system $\overline{\mathscr{A}}_{K+1}$ and, 2) the (PS) per-customer service rate is lower between consecutive events in system $\overline{\mathscr{A}}_{K+1}$, compared to $\mathscr{A}_K(\lambda_K)$.

C) From the above results, we deduce that event $(\overline{\mathbf{N}}(t) = \mathbf{0})$ implies event $(\mathbf{N}(t) = \mathbf{0})$ for all $t \geq 0$. In the stationary regime, we derive that $(\overline{\mathbf{N}} = \mathbf{0}) \subset (\mathbf{N} = \mathbf{0})$ and thus $Q(\lambda_K') \leq Q(\lambda_K)$. We conclude that $\lambda_K \mapsto Q(\lambda_K)$ is a decreasing function.

D) From the same inclusion argument, we deduce that

$$\mathbb{P}(\mathbf{N} = \mathbf{0}) = \mathbb{P}(\overline{\mathbf{N}} = \mathbf{0}) + \mathbb{P}(\mathbf{N} = \mathbf{0},\, \overline{\mathbf{N}} \neq \mathbf{0}). \tag{9}$$

Noting that $(\overline{N}_{K+1} > 0;\, \forall\, k \in \{1, ..., K\}, \overline{N}_k = 0) \subset (\overline{\mathbf{N}} \neq \mathbf{0};\, \forall\, k \in \{1, ..., K\}, N_k = 0)$ where the inclusion follows by the property derived in **B)**, we deduce that

$$\mathbb{P}(\mathbf{N} = \mathbf{0},\, \overline{\mathbf{N}} \neq \mathbf{0}) \geq \mathbb{P}(\overline{N}_{K+1} > 0;\, \forall\, k \in \{1, ..., K\}, \overline{N}_k = 0) > 0, \tag{10}$$

since the distribution of $\overline{\mathbf{N}}$ gives positive weight to any subset of its range. After (8), (9) and (10), we derive the strict decreasing behaviour of $Q(.)$, as claimed. □

Proposition 2.4 ***Function $Q(.)$ is continuous over its definition interval.***

Proof. The derivation proceeds according to the following steps.

A) Keeping the same notation, the right-continuity of $Q(.)$ at a given point λ_K will follow if it is shown that $\mathbb{P}(\mathbf{N}=\mathbf{0},\,\overline{\mathbf{N}}\neq\mathbf{0})$ tends to 0 when $\Delta\lambda_K$ tends to 0.

Consider the joint Markov process $(\mathbf{N}(t),\overline{\mathbf{N}}(t))_{t\geq 0}$, starting with empty queues, and a cycle of given duration $\overline{\tau}$ starting at $t=0$, without loss of generality. First note, in view of property **B)** above, that a cycle of this joint process is identical to a cycle of process $(\overline{\mathbf{N}}(t))_{t\geq 0}$. Then the event $(\mathbf{N}(t)=\mathbf{0},\ \overline{\mathbf{N}}(t)\neq\mathbf{0})$ for given $t\in[0,\overline{\tau}]$ implies that the date τ_0 of the first arrival from class $(K+1)$ is no greater than t, that is, $\mathbf{1}_{(\mathbf{N}(t)=\mathbf{0},\,\overline{\mathbf{N}}(t)\neq\mathbf{0})}\leq\mathbf{1}_{t\geq\tau_0}$. We thus derive that

$$\mathbb{E}\left(\int_0^{\overline{\tau}}\mathbf{1}_{(\mathbf{N}(t)=\mathbf{0},\,\overline{\mathbf{N}}(t)\neq\mathbf{0})}\,\mathrm{d}t\right)\leq\mathbb{E}\left(\int_0^{\overline{\tau}}\mathbf{1}_{(t\geq\tau_0)}\,\mathrm{d}t\right)\leq\mathbb{E}\left(\overline{\tau}\cdot\mathbf{1}_{(0\leq\tau_0\leq\overline{\tau})}\right).$$

We now apply the cycle formula ([3], Chap. IV, Theorem 8.4) to the ergodic process $(\mathbf{N}(t),\overline{\mathbf{N}}(t))_{t\geq 0}$ and use the Cauchy-Schwarz inequality to get

$$\mathbb{P}(\mathbf{N}=\mathbf{0},\,\overline{\mathbf{N}}\neq\mathbf{0})=\frac{1}{\mathbb{E}(\overline{\tau})}\cdot\mathbb{E}\left(\int_0^{\overline{\tau}}\mathbf{1}_{(\mathbf{N}(t)=\mathbf{0},\,\overline{\mathbf{N}}(t)\neq\mathbf{0})}\,\mathrm{d}t\right)\leq\frac{\mathbb{E}(\overline{\tau}^2)^{1/2}}{\mathbb{E}(\overline{\tau})}\cdot\mathbb{P}(0\leq\tau_0\leq\overline{\tau})^{1/2}.$$

B) Decomposing the cycle duration $\overline{\tau}$ as the sum $\overline{\tau}=\overline{\tau}_B+\overline{\tau}_I$ of a busy period duration $\overline{\tau}_B$ and of the following idle period duration $\overline{\tau}_I$, it is shown that the first and second moments of $\overline{\tau}$ are locally (i. e., around $\Delta\lambda_K=0$) lower- and upper-bounded, respectively. First, $\overline{\tau}_I$ is exponentially distributed with parameter $\Lambda+\Delta\lambda_K$ (where $\Lambda=\lambda_1+...+\lambda_K$), so that $\mathbb{E}(\overline{\tau})\geq\mathbb{E}(\overline{\tau_I})=1/(\Lambda+\Delta\lambda_K)$. Second, noting that random variables $\overline{\tau}_B$ and $\overline{\tau}_I$ are independent (due to the Poisson arrival processes), we have $\mathbb{E}(\overline{\tau}^2)=\mathbb{E}(\overline{\tau_B}^2)+2\mathbb{E}(\overline{\tau_B})\,\mathbb{E}(\overline{\tau_I})+\mathbb{E}(\overline{\tau_I}^2)$.

To ensure that $\mathbb{E}(\overline{\tau}_B)$ and $\mathbb{E}(\overline{\tau}_B^2)$ are finite and locally bounded functions of $\Delta\lambda_K$, we state that they are upper-bounded by the corresponding moments of τ_B^*, the busy period of the same system without impatience. Then, by gathering all $K+1$ classes into a single class, τ_B^* is also the busy period of a one-class $M/G/1$ PS queue with a compound distribution for the data volume $\overline{B}$. Since the system without impatience is work-conserving, the distribution of τ_B^* is independent of the actual service discipline; its first and second moments are continuous functions of the load and moments of $\overline{B}$, as shown by formulas given in ([12], Vol.I, Chap.5, Section 5.8, Equ.(5.141) and Equ.(5.142)). As a consequence, the moments of $\overline{\tau}_B$, and thus $\mathbb{E}(\overline{\tau}^2)$, are locally upper-bounded.

C) It remains to show that $\mathbb{P}(0\leq\tau_0\leq\overline{\tau})$ tends to 0. For any $A>0$, write

$$\mathbb{P}(\tau_0\leq\overline{\tau})=\mathbb{P}(\tau_0\leq\overline{\tau},\tau_0\leq A)+\mathbb{P}(A\leq\tau_0\leq\overline{\tau})\leq\mathbb{P}(\tau_0\leq A)+\mathbb{P}(A\leq\overline{\tau}).\quad(11)$$

By the Markov inequality, we first have $\mathbb{P}(A\leq\overline{\tau})\leq\mathbb{E}(\overline{\tau}^2)/A^2$ so that, since $\mathbb{E}(\overline{\tau}^2)$ is locally bounded, we can select A such that $\mathbb{P}(A\leq\overline{\tau})$ is arbitrarily small, say, lower than ε for any given $\varepsilon > 0$. For this value of A, let now $\Delta\lambda_K$ tend to 0. The distribution of τ_0 is independent from the cycle duration and is a compound of an atom at 0 and an exponential distribution, hence $\mathbb{P}(\tau_0 \leq A) = \alpha + (1-\alpha)(1-e^{-A\Delta\lambda_K})$ with $\alpha = \Delta\lambda_K/(\Lambda + \Delta\lambda_K)$. This probability can thus be made lower than ε for small enough $\Delta\lambda_K$, thus making $\mathbb{P}(\tau_0 \leq \overline{\tau})$ lower than 2ε after (11). This finally justifies the right-continuity of function Q at point λ_K.

D) A similar reasoning shows that function Q is also left-continuous. □

3 Network with Inter-cell Mobility

We now address the description of a whole network of cells where users move from one cell to neighboring ones due to possible handovers.

3.1 A Closed Network of Queues

Consider a cellular network of I cells with possibly distinct capacities. Users from K traffic classes may appear and move during their communications. When leaving a cell during transmission, users join one of the neighboring cells according to some routing probabilities. They consequently generate supplementary flows of new arrivals, hereafter called *handover arrivals*, which are to be added to *fresh arrivals* in each cell. Assume that class-k users generate requests for transmission in cell i according to a Poisson process with rate $\lambda^0_{i,k}$, $i = 1,...,I$, $k = 1,...,K$; this corresponds to the fresh traffic offered to cell i. To account for class-k users that became active outside cell i and experienced some handovers, the total flow arrival to cell i is

$$\lambda_{i,k} = \lambda^0_{i,k} + \lambda^{In}_{i,k} = \lambda^0_{i,k} + \sum_{j\neq i} p_k(j,i)\cdot\lambda^{Out}_{j,k}, \tag{12}$$

where $\lambda^{In}_{i,k}$ denotes the handover arrival rate from neighboring cells, $\lambda^{Out}_{j,k}$ is the handover departure rate from cell j, and $p_k(j,i)$ are the routing probabilities from cell $j \neq i$ to cell i. For all i and k, we will assume that the handover arrival process to cell i from class-k users can be approximated by a Poisson process so that it can be superposed to the fresh arrivals to build up a total Poisson arrival process with rate $\lambda_{i,k}$ given in (12). All Poisson processes introduced above are supposed to be mutually independent, an assumption which notably simplifies the global description of the system by reducing it to a network of queues which is *closed* regarding the handover flows. This is, in particular, in contrast to the overall multi-class multi-cell process considered in other papers [4, 5, 7, 11].

The rate $\lambda^{Out}_{j,k}$ can in turn be considered as an output of the generic queuing model considered in Section 2 for cell j, and be calculated by means of some *performance function* $\mathscr{F}_{j,k}(.)$, that is,

$$\lambda_{j,k}^{Out} = \mathscr{F}_{j,k}\left(\lambda_{j,1}^{In},...,\lambda_{j,K}^{In}\right) = \theta_{j,k}\mathbb{E}(N_{j,k}) \tag{13}$$

for $j = 1,...,I$ and $k = 1,...,K$, where $\theta_{j,k}$ (resp. $N_{j,k}$) denotes the mobility rate from cell j emanating from class-k ongoing transfers (resp. the number of class-k ongoing transfers in cell j). In (13), only handover arrival rates are considered as variables, all other intrinsic parameters (such as cell capacities, per-class offered traffic and mobility rates) being kept constant. From (12)-(13), it follows that a stationary network regime can be characterized by a system of $I \times K$ flow equations with the handover arrival rates $\lambda_{i,k}^{In}$ as unknowns, namely

$$\lambda_{i,k}^{In} = \sum_{j \neq i} p_k(j,i) \cdot \mathscr{F}_{j,k}\left(\lambda_{j,1}^{In},...,\lambda_{j,K}^{In}\right). \tag{14}$$

The problem of existence and uniqueness of a solution to the non-linear system (14) is out of the scope of the present paper. As the performance functions $\mathscr{F}_{j,k}$ may not be explicit in terms of input parameters, the practical determination of a solution to (14) involves a numerical iterative procedure, e.g. a fixed-point algorithm.

3.2 The case of a homogeneous network

Now assume that the network is *homogeneous* in the following sense:

(i) all intrinsic parameters (capacities, arrival rates, ...) are the same for all cells, so that performance functions do not depend on the cell, that is, $\mathscr{F}_{j,k}(.) = \mathscr{F}_k(.)$;

(ii) the routing of handover flows is symmetric, i.e., for each class k, cell i receives handover traffic from a set $\mathscr{J}_k(i)$ of neighboring cells with identical probability $p_k(j,i) = 1/J_k$, where J_k is the common cardinal of sets $\mathscr{J}_k(i)$.

Clearly, any set of rates $\lambda_{i,k}^{In} = \lambda_k^{In}, \forall\, i,k$, verifying the simpler system

$$\forall k \in \{1,...,K\}, \quad \lambda_k^{In} = \lambda_k^{Out} = \mathscr{F}_k\left(\lambda_1^{In},...,\lambda_K^{In}\right), \tag{15}$$

will provide a particular solution to (14), hence *the* solution if uniqueness is ensured. For a homogeneous network, the problem thus reduces to the study of a single *representative cell*, where the ingress and outgoing handover traffics balance exactly.

In this context, we define the total and per-class loads by referring to the fresh traffic, that is, $\rho_k^0 = \lambda_k^0 \sigma_k / C, 1 \leq k \leq K$, and $\rho^0 = \sum_{k=1}^{K} \rho_k^0$. We now claim that the equilibrium of this system is characterized by $\rho^0 < 1$, a condition which is well understood since mobile users re-enter the system until their transfer is completed.

Proposition 3.1 ***A) In the homogeneous network with inter-cell mobility,***

$$\rho^0 < 1 \tag{16}$$

is a necessary condition for the existence and uniqueness of a fixed-point solution to equilibrium equations (15), that is, $\lambda_k^{In} = \lambda_k^{Out}, 1 \leq k \leq K$.

B) In the specific case of a single mobile class, this condition is also sufficient.

Proof. **A)** Assume that there exists a solution to equilibrium equations (15). For the system associated with that solution, we apply (2) to any class k; recalling that $\lambda_k = \lambda_k^0 + \lambda_k^{In}$ by (12) and that $\lambda_k^{In} = \lambda_k^{Out} = \theta_k \mathbb{E}(N_k)$ by (13), we then obtain $\rho_k^0 = \mathbb{E}\left(N_k \mathbf{1}_{N_k>0}/L(\mathbf{N})\right)$, $1 \leq k \leq K$. The side-by-side summation of these equalities yields the required condition $\rho^0 < 1$, after observing that

$$\sum_{k=1}^{K} \mathbb{E}\left(\frac{N_k \mathbf{1}_{N_k>0}}{L(\mathbf{N})}\right) = \mathbb{E}(\mathbf{1}_{\mathbf{N}\neq(\mathbf{0})}) = 1 - \mathbb{P}(\mathbf{N} = \mathbf{0}). \tag{17}$$

B) To address the sufficiency of (16), first note that $\rho_S = \rho_S^0 < 1$ obviously holds, ensuring that the queue is stable whatever the load of mobile users. Besides, writing relation (2) for any class k yields $\lambda_k^0 + \lambda_k^{In} = \mu_k \mathbb{E}\left(N_k \mathbf{1}_{N_k>0}/L(\mathbf{N})\right) + \lambda_k^{Out}$ which, after dividing each side by μ_k and summing over all $k \in \{1, ..., K\}$, gives

$$\rho^0 = \sum_{k=1}^{K} \frac{\lambda_k^0}{\mu_k} = 1 - \mathbb{P}(\mathbf{N} = \mathbf{0}) + \sum_{k=1}^{K} \frac{\lambda_k^{Out} - \lambda_k^{In}}{\mu_k}. \tag{18}$$

At this stage, we assume that there is only one single class M of mobile users. As $\lambda_k^{In} = \lambda_k^{Out} = 0$ for all static classes k, it follows from (18) that the left equilibrium equation in (15), $\lambda_M^{In} = \lambda_M^{Out}$, is equivalent to

$$\mathbb{P}(\mathbf{N} = \mathbf{0}) = 1 - \rho^0. \tag{19}$$

We use the notation of Section 2.2 to class M so that $\mathbb{P}(\mathbf{N} = \mathbf{0}) = Q(\lambda_M)$ is a function of the arrival rate $\lambda_M = \lambda_M^0 + \lambda_M^{In}$ of class-M users. Since $\rho^0 < 1$ is assumed, the existence of a unique solution λ_M to (19) is ensured if it is shown that

I. $Q(\lambda_M^0) \geq 1 - \rho^0$;
II. $Q(.)$ is strictly decreasing over $\mathbb{R}^+$;
III. $Q(.)$ is continuous over $\mathbb{R}^+$;
IV. $Q(\lambda) \to 0$ as $\lambda \to +\infty$.

Items **I**, **II**, **III** and **IV** can be successively proved as follows.

I. For any impatience queuing system, summing the conservation relation (2) applied to each class and recalling (17), we get the following expression for the average number of moving users:

$$\mathbb{E}(N_M) = \frac{\mu_M}{\theta_M}\left(\rho + \mathbb{P}(\mathbf{N} = \mathbf{0}) - 1\right). \tag{20}$$

By identity (20) for the mean number $\mathbb{E}(N_M)$, its non-negativity entails that

$$\forall\, \lambda_M \geq 0, \quad Q(\lambda_M) = \mathbb{P}(\mathbf{N} = \mathbf{0}) \geq 1 - \rho. \tag{21}$$

Lower bound (21) holds, in particular, for the value $\lambda_M = \lambda_M^0$ for which $\rho = \rho^0$, hence $Q(\lambda_M^0) \geq 1 - \rho^0$ as required.

II. This is ensured by Proposition 2.3.

III. This is ensured by Proposition 2.4.

IV. All parameters being kept constant, we denote by N_M^* the number of mobile users in the same queuing system but in the absence of any static users. A straightforward sample path argument enables us to show that $N_M \geq N_M^*$ almost surely, so that

$$Q(\lambda_M) = \mathbb{P}(\mathbf{N} = \mathbf{0}) \leq \mathbb{P}(N_M = 0) \leq \mathbb{P}(N_M^* = 0). \tag{22}$$

Besides, the stationary distribution of N_M^* is that of the occupancy process of a PS queue with a single customer class M [8], with load $\rho_M = \lambda_M/\mu_M$ and impatience rate θ_M; the corresponding empty-queue probability is, in particular, given by

$$\mathbb{P}(N_M^* = 0) = \left(1 + \sum_{\ell \geq 1} \rho_M^\ell \left[\prod_{1 \leq j \leq \ell} (1 + j\theta_M/\mu_M)\right]^{-1}\right)^{-1}.$$

This probability is then upper-bounded by $(1 + \theta_M/\mu_M)/\rho_M$, which ensures together with inequality (22), that $Q(\lambda_M)$ tends to 0 as λ_M tends to infinity. □

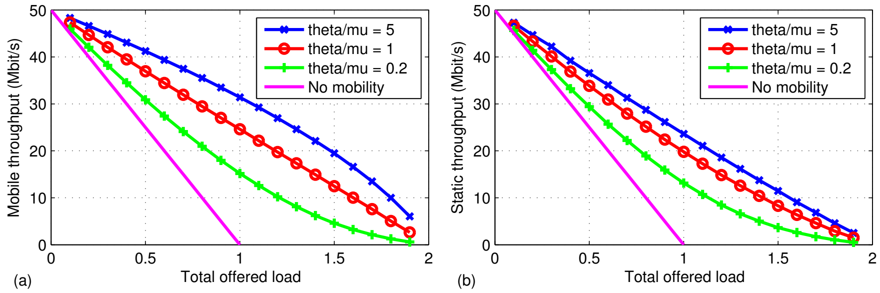
4 Numerical results

We report here some numerical experiments, focusing on the important case where users are gathered into two classes, namely one static and one mobile class. In all subsequent scenarios, we fix a cell capacity $C = 50$ Mbit/s, a proportion of 50% mobile users and a mean flow volume $\sigma = 12.5$ MB (100 Mbit) for both classes.

4.1 Impatience Model

Regarding the generic Markovian model analyzed in Section 2, we examined in [14] the sensitivity of the performance indicators to the distributions of sojourn (impatience) time and flow volumes. It was shown there that the throughput of each class is only marginally impacted by both distributions, indicating that results derived from the Markovian framework remain valid for more realistic distributions, while the handover probability is noticeably more impacted (particularly at low load), increases with the variance of T_M and decreases with the variance of Σ_M and Σ_S.

We now focus on the performance indicators provided by the (numerically solved) Impatience Model. In Fig. 1 are plotted the average throughputs and handover probability obtained with exponentially distributed T_M, Σ_M and Σ_S, and considering a series of three normalized mobility rates: θ_M equals 0.2, 1 or 5 times the service capacity $\mu_M = 0.5$ s^{-1}. Large throughput gains for static and mobile users are observed, compared to a scenario where all users would be static; besides, we note a significant gain of mobile users throughput over that of static users. We have further observed that the throughput gain of mobile over static users appears to be



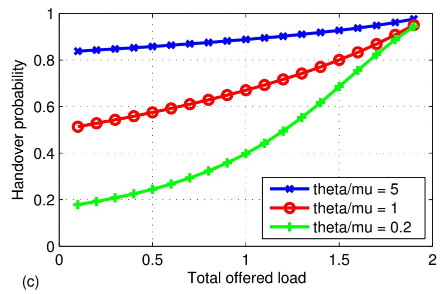


Fig. 1 (a) mobile throughput, (b) static throughput and (c) handover probability obtained from the Impatience Model, for 50% mobile users and a normalized mobility rate of 0.2, 1, or 5.

the greatest when the mobility rate is large (say, more than twice the service rate for mobile users) and when the proportion of mobile users is small (say, 20%).

Such throughput gains are expected in this open-loop system, and were already observed in [2]: they result from the impatient nature of mobile users who may leave the system without re-entering it (hence the gain w.r.t. the all-static scenario) and tend to do it especially when local congestion occurs (hence the gain mobile/static). Very large handover rates, see Fig. 1c), may however counterbalance these gains.

Interpreting the above results helps us to assess the impact of cell size. Assuming a constant speed v, the mean distance the mobile user travels in the cell is $\mathbb{E}(D) = v/\theta_M$; this mean distance is typically of the order of the radius R of a circular cell. Thus if $v = 90$ km/h for example, the values of θ_M considered above, namely $5\,\mu_M$, μ_M and $0.2\,\mu_M$ (with $\mu_M = 0.5$ s^{-1}) respectively correspond to a radius of 10 m, 50 m and 250 m, typical of a Femto, Pico, and Micro cell. As expected, users in Femto cells experience the largest throughput since their mobility rate is the highest.

4.2 Mobility Model

We assess the Mobility Model proposed in Section 3, i.e., the Markovian model where the handover arrival rate λ_M^{In} exactly balances the outgoing handover rate λ_M^{Out}. We consider the homogeneous four-cell ring network shown in Fig. 2, where

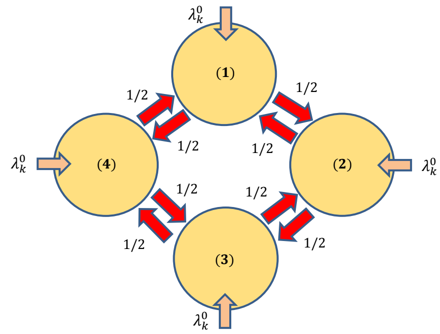


Fig. 2 A homogeneous ring network of four identical cells with symmetric routing.

all cells have the same capacity and traffic parameters as in Section 4.1, and three normalized mobility rates: θ_M is equal to 0.1, 1 or 10 times the service capacity μ_M.

Event-driven simulations have been performed at the flow level. The accuracy of results drawn from simulation has been tightly controlled. In every configuration, ten independent simulation runs have been performed, generating around 1 million discrete events each. The obtained confidence intervals are very small in most cases and thus are not shown in the following plots for simplicity.

Fig. 3 depicts the performance indicators for mobile users in each cell. For each value of θ_M, the four curves (each corresponding to one cell) are almost indistinguishable from each other, thus assessing the robustness of simulations. We observe that the stability region is characterized by $\rho^0 < 1$, as predicted by Proposition 3.1 and, from Fig. 3(a), that the throughput gains due to mobility increase with the mobility rate. Other complementary results have shown that the mobile/static throughput gain is all the more important that the proportion of mobile users is weak. The latter simulation is compared in Fig. 4 to the Mobility Model (applied to the *representative cell*) when $\theta_M = \mu_M$. We observe that the representative cell model provides slightly optimistic throughputs compared to those obtained from simulation.

The good match between model and simulation results validates our approach for reducing a homogeneous network to a single representative cell: the assumption quoted in Section 3.1, that the handover traffic flow re-enters the representative cell as a supplementary Poisson flow, appears reasonable. The robustness of the latter assumption has also been checked in the case of a heterogeneous ring network [14].

Finally, we evaluate the impact of mobile speed for a given cell size (a cell radius of 50 m corresponding to a Pico cell). Fig. 5 depicts the static and mobile users throughputs in terms of the speed v for different values of the total offered load (0.2, 0.5, and 0.8). Results are here derived from the Markovian Mobility Model only. As expected, all performance indicators are increasing functions of the speed; but note that the impact at very high speed is rather limited.

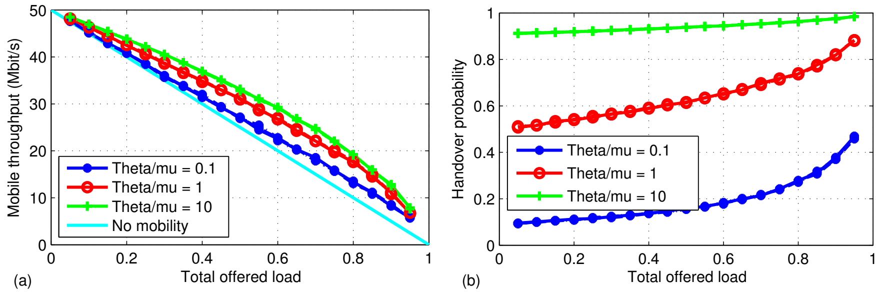


Fig. 3 Homogeneous ring network: performance indicators, (a) mobile user throughput and (b) handover probability, obtained from simulation versus the total offered load in each cell (proportion of 50% mobile users and $\theta_M/\mu_M = 0.1, 1, 10$).

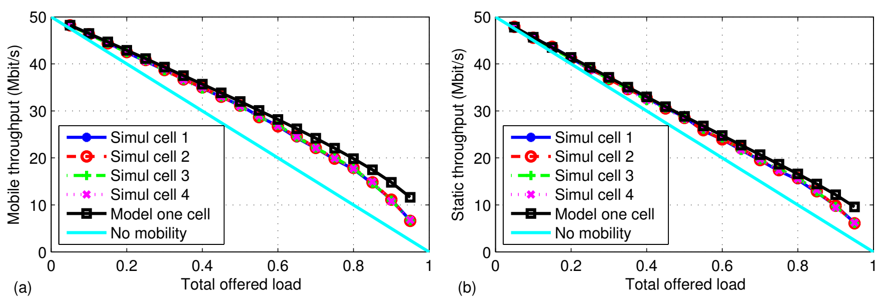


Fig. 4 Homogeneous ring network: (a) mobile throughput and (b) static throughput obtained from simulation and the Markovian Mobility Model ($\theta_M/\mu_M = 1$).

5 Conclusion

We have investigated the impact of inter-cell mobility on the performance of dense networks with small cells. Our approach relies on two main ideas: a simple performance model can be developed to capture mobility on the basis of the multi-class Processor-Sharing queue with impatience; the performance of a network of small cells can be handled by applying the generic model to each individual cell. The present paper extends the former contribution [14], notably by providing mathematical proofs for the stability of the impatience model and for the existence and uniqueness of an equilibrium regime for the handover flows in a homogeneous network. Further practical outcomes can be stated as follows: **(i)** as a step beyond available studies, the handover probability has been evaluated to assess the trade-off between throughput gain and signaling overhead due to mobility; **(ii)** both classes of users are shown to benefit from a throughput gain induced by inter-cell mobility; this gain is created by the opportunistic displacement of mobile users within the network according to local load variations in individual cells.

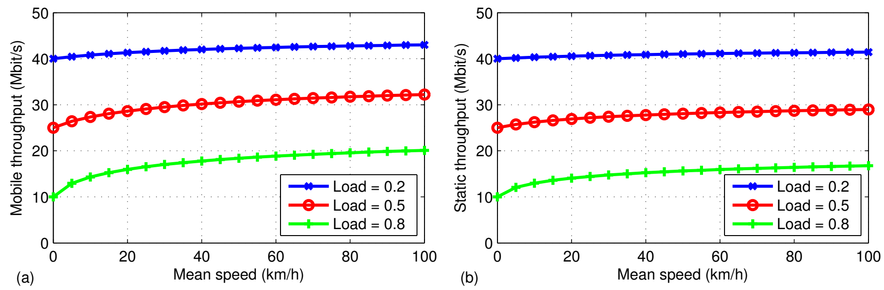


Fig. 5 Impact of the users speed on (a) mobile throughput and (b) static throughput for a proportion of 50% mobile users, cell radius 50 m, and total offered load 0.2, 0.5 or 0.8.

References

1. N. Abbas, T. Bonald, and B. Sayrac. Opportunistic Gains of Mobility in Cellular Data Networks. In *Proc. of WiOpt 13th Intl. Symposium*, 2015.
2. B. Baynat, R.-M. Indre, N. Nya, P. Olivier, and A. Simonian. Impact of Mobility in Dense LTE-A Networks with Small Cells. In *Proc. of VTC2015- Spring Workshop*, Glasgow, Scotland, May 2015.
3. R. Bhattacharya and E. Waymire. *Stochastic Processes with Applications*, volume 61 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), 2009.
4. T. Bonald, S. Borst, and N. Hegde. How Mobility Impacts the Flow-level Performance of Wireless Data Systems. In *Proc. of IEEE Infocom*, 2004.
5. T. Bonald, S. Borst, N. Hegde, M. Jonckheere, and A. Proutière. Flow-level Performance and Capacity of Wireless Networks with User Mobility. *Queueing Systems*, 63:131–164, 2009.
6. T. Bonald and A. Proutière. Wireless Downlink Data Channels: User Performance and Cell Dimensioning. In *Proc. of MobiCom'03*, San Diego, USA, Sept. 2003.
7. S. Borst, A. Proutière, and N. Hegde. Capacity of Wireless Data Networks with Intra- and Inter-cell Mobility. In *Proc. of IEEE Infocom*, 2006.
8. E. G. Coffman, A. A. Puhalskii, M. I. Reiman, and P. E. Wright. Processor-Shared Buffers with Reneging. *Performance Evaluation*, 19:25–46, 1994.
9. E. Dahlman, S. Parkvall, and J. Skold. *4G: LTE/ LTE-Advanced for Mobile Broadband (2nd ed.)*. Academic Press, 2013.
10. F. Guillemin, S. ElAyoubi, C. Fricker, P. Robert, and B. Sericola. Controlling Impatience in Cellular Networks Using QoE-aware Radio Resource Allocation. In *Proc. of 27th International Teletraffic Congress*, 2015.
11. M. K. Karray. User's Mobility Effect on the Performance of Wireless Cellular Networks Serving Elastic Traffic. *Wireless Networks*, 17:247–262, 2011.
12. L. Kleinrock. *Queueing Systems*, volume I: Theory. Wiley Interscience, 1975.
13. X. Lin, R. K. Ganti, P. J. Fleming, and J. G. Andrews. Towards Understanding the Fundamentals of Mobility in Cellular Networks. *IEEE Transactions on Wireless Communications*, 12(4), April 2013.
14. P. Olivier and A. Simonian. Performance of Data Traffic in Small Cells Networks with Inter-Cell Mobility. In *Proc. of ValueTools 2016*, Taormina, Italy, Oct. 2016.
15. A. Osseiran et al. Scenarios for 5G Mobile and Wireless Communications: the Vision of the METIS Project. *IEEE Communications Magazine*, 52(5):26–35, 2014.
16. P. Robert. *Stochastic networks and queues*. Springer, 2003.
17. C. Schindelhauer. Mobility in Wireless Networks. In *Proc. of SOFSEM 2006: Theory and Practice of Computer Science*, Merin, Czech Republic, Jan. 2006.