Performance Evaluation of Scalable Congestion Control Schemes for Elastic Traffic in Cellular Networks with Power Control

Bartłomiej Błaszczyszyn[1] and Mohamed Kadhem Karray[2]

Abstract— **This paper deals with the performance evaluation of some congestion control schemes for elastic traffic in wireless cellular networks with power allocation/control. These schemes allow us to identify the *feasible configurations of instantaneous* up- and downlink *bit-rates* of users; i.e., such that *can be obtained by allocating respective powers*, taking into account in an exact way the interference created in the whole, multicellular network. We consider the bit-rate configurations identified by these schemes as *feasible sets for some classical, maximal fair resource allocation policies*, and study their performance in the long-term evolution of the system. Specifically, we assume Markovian arrivals, departures and mobility of customers, which transmit some given data-volumes, as well as some temporal channel variability (fading), and study the *mean number of users*, the *mean throughput* i.e., the mean bit-rates, and the *mean delay* that these policies offer in different parts of a given cell. Explicit formulas are obtained in the case of *proportional fair policies*, which may or may-not take advantage of the fading, for *null* or *infinitely rapid customer mobility*. This approach applies also to a channel shared by the elastic traffic and a streaming, with predefined customer bit-rates, regulated by the respective admission policy.**

I. INTRODUCTION

Cellular networks provide real-time services (voice calls, on-line streaming), which require predefined transmission rates, and carry elastic traffic (data), which accepts fluctuations of the rates. In order to use the shared medium more efficiently, the power control is often used (as e.g in CDMA or WIMAX networks). In the interference limited networks it is the whole configuration of powers allocated to different users that creates the interference and thus determines the actual transmission rates.

Growing traffic requires better planning and/or dimensioning of the networks. This task can be substantially simplified by analytical study of the network performance. One distinguishes two milestones of the analytical evaluation of the network performance: identification of its stability region, and the evaluation of the steady state characteristics (e.g. the mean throughput).

The following two groups of elements are crucial for the network performance. On one hand, the network geometry, the coding and the service policy (bit-rate assignment), which can be chosen by the network designer. On the other hand, channel conditions and the user traffic (intensity of arrivals, mobility and transfered data volumes) can only be predicted by the designer. In order to optimize the network performance, the coding and the service policy should be adapted to the channel and traffic conditions. Besides better coding (up to the Shannon's limit) one is looking for policies that are opportunistic; i.e., take advantage of the actual channel state and user position, preserving some fairness of the mean service rates offered to users. Given the optimal available choices of coding and policy, the designer has to find a network geometry (e.g. cell size of the regular hexagonal pattern) that can cope with the given traffic under some quality of service constraints. Individual elements of the above puzzle are often studied and optimized separately (see Section II). In particular the geometry of interference is often absent or seriously simplified.

The general objective of this paper is to propose a global approach to the performance evaluation of cellular networks with elastic traffic, which combines the existing detailed results, in particular concerning the performance of the opportunistic fair policies and the impact of the mobility, with our previous results on the geometry of interferences in power controlled cellular networks.

We propose a model that assumes Markovian arrivals, departures and mobility of customers, which transmit some given data-volumes, as well as some temporal channel variability (fading). Fair transmission rate assignment policies are deduced from schemes, which ensure globally in the network existence of a solution of the respective power control problem. We study the mean throughput i.e., the mean bit-rates that these policies offer in different parts of a given cells.

The main results of this paper are explicit formulas for the mean number of users, the mean throughput and the mean delay obtained in the case of proportional fair policies, which may or may-not take advantage of the fading, for null or infinitely rapid customer mobility. These results are obtained by an appropriate separation of the times scales of different elements of the network dynamics.

We apply our results to the analysis of the Downlink Shared Channel (DSCH) of the Universal Mobile Telecommunications System (UMTS) release 99, which utilizes a non-opportunistic service policy, and the UMTS release High Speed Downlink Packet Access (HSDPA) — the 3GPP analogue of the High Data Rate (HDR) channel in the Code Division Multiple Access CDMA2000 1xEV-DO standard, which relies on a channel-aware policy.

The remaining part of this paper is organized as follows. In Section II we list the results in the literature that are used in our global approach. In Section III we recall briefly the idea of the decentralized congestion control schemes for power-controlled networks developed in [1], [2]. The fair instantaneous bit-rates offered by these schemes are discussed in Section IV. The network dynamics is described in Section V. We describe there also the principle of the time scale separation. The main results of the performance evaluation are given in Section VI and followed by some numerical examples in Section VII. In the Appendix we introduce and study a mathematical model for wireless networks serving elastic traffic (WET) that is used in the paper for the performance evaluation of the considered schemes and policies.

[1] INRIA & Ecole Normale Supérieure and Mathematical Institute University of Wrocław, 45 rue d'Ulm, 75005 Paris France, Bartek.Blaszczyszyn@ens.fr,

[2] France Télécom R&D, 38/40 rue du Général Leclerc, 92794 Issy-Moulineaux France, mohamed.karray@orange.ft.com

II. RELATED WORK

There is a rich literature on wireless data networks. It is beyond the scope of this paper to make a thorough review of it; for this we refer e.g. to a recent paper [3]. Below, we only review those results that are directly related to our approach.

Our WET model, used for the performance evaluation of the considered schemes and policies assumes Markovian arrivals, departures and mobility of customers, which transmit some given data-volumes. It can be seen as an extension of the Spatial Queueing System described in [4, Chapter 10], which allows for departure rates depending, not only on the current, but also on the next location of a given user (besides current locations of other users).

This model is different form the Whittle model proposed for such networks in [3, § 2]. The fundamental difference is that in the WET model each user has some given data-volume to transmit during the whole sojourn in the system, while in the Whittle model of [3, § 2] each user has a different data-volume to transmit at each visited location, and he does not move from this location until the end of the transmission of the required volume. Consequently, when the Whittle model approaches a congestion, user mobility is being frozen, whereas it is not influenced by a congestion in the WET model.

In order to evaluate explicitly the impact of user mobility on the performance of the system, we follow the idea presented in [5] of quasi-stationary and fluid limit, and consider two extreme cases of motionless and infinitely-rapid users in the WET model. In these cases and a proportional fair policy, the WET model becomes equivalent to a spatial extension or, respectively, the classical version of the processor sharing queue, and we rely on the well know results for its stability and stationary distribution; see e.g. [6], [4], [7]. These results are enough to analyze the case without macrodiversity; i.e., when each user is served by exactly one base station. More general results on stability of wireless data networks can be found in [8].

In order to study the channel-aware rate-assignment policies we enrich the WET model with independent fluctuations of the channel conditions of each user, and follow [9] that gives the performance of an opportunistic weight-based rate allocation policy. This policy is shown in [10] to be a good model for opportunistic schedulers implemented in HDR and HSDPA channels. However, we consider also a non-opportunistic proportional fair policy that we propose as a model e.g. for the DSCH channel. The novelty of our approach is that, when identifying the feasible bit-rate allocations, we build upon the scalable schemes developed in [1], [2]. These schemes are decentralized, in that they can be implemented in such a way that each base station only has to consider the load brought by its own users. Moreover, they allow one to control in one channel the elastic traffic and a streaming traffic, with predefined customer bit-rates, regulated by an appropriate admission policy (see [11] for the study of the respective blocking rates). Our approach applies to large, multi-cell networks, and it is different from [12], as it is based on the sub-stochasticity condition imposed on some matrix of relative path-losses, introduced in [13], whose spectral radius when less than one is equivalent to the global feasibility of the power allocation.

III. DECENTRALIZED CONGESTION CONTROL

In this section we recall briefly the decentralized power and congestion control schemes for cellular networks developed in [1], [2]. Our canonical example is of the downlink in a CDMA network, but the approach applies to both up- and downlink of other cellular networks with power control. The goal consists in finding *decentralized constraints on the SINR threshold's* (and hence, instantaneous bit-rates) which guarantee the global feasibility of the power control in the network.

A. Notation

We will use the following notation:

1) Antenna Locations and Path-Loss: $\{Y^u\}_u$ denotes the locations of BS's; Y^u denotes the location of the BS with index u; S_u is the set of mobiles served by BS u; $\{X^u_m\}_m$, with $m \in S_u$, denotes the locations of the mobiles served by BS u; l^v_m is the path-loss $Y^v \to X_m$;

2) Engineering Parameters: α_u denotes the downlink (DL) orthogonality factor in BS u; let $\alpha_{uv} = \alpha_u$ if $u = v$ and 1 otherwise. P^u_m is the power of the dedicated channel $u \to m$; N^u_m is the external noise at mobile $m \in S_u$; ξ^u_m is the SINR threshold for user X^u_m. It is *related to the bit-rate* of that user. For each SINR ξ, we define also a *modified SINR* $\xi^{'}$ by

$$\xi'^u_m = \frac{\xi^u_m}{1+\alpha_u\xi^u_m}\,, \tag{3.1}$$

3) Bit-Rates: We denote by R, R^u_m bit-rates. For a particular coding and modulation scheme used in the channel, and given a required link quality (bit-error rate), there is one-to-one correspondence between the attained SINR threshold ξ and the bit-rate R that can be reliably sustained in this channel. In particluar, in the Dedicated Channels (DCH) and Downlink Shared Channels (DSCH) of the UMTS release 99 the following relation is implemented

$$R = \frac{\xi W}{(E_b/I_0)}\,, \tag{3.2}$$

where (E_b/I_0) is the bit-energy-to-noise-density ratio (related to the required link quality), W is the chip-rate. We remind that the *theoretical maximal bit-rate* of the Gaussian channel (AWGN) is related to the SINR ξ by the Shannon's formula

$$R = B\log_2(1+\xi)\,, \tag{3.3}$$

where B is the bandwidth. [3]

B. Power Control

Taking the downlink as an example, we now recall the so called *feasibility problem* of the power control (equivalently, power allocation); see [2], [11] for the formulation of the both up- and downlink problems with power constraints.

We will say that the *power allocation is feasible* if there exist nonnegative and finite powers P^u_m, for all base stations u and all mobiles $m \in S_u$, which satisfy the following condition: signal to interference and noise ratio at each mobile is larger than the threshold ξ^u_m, i.e.

$$\frac{P^u_m/l^u_m}{N^u_m + \sum_v \alpha_{uv}(\sum_{n\in S_v} P^v_n)/l^v_m} \ge \xi'^u_m\,, \tag{3.4}$$

for all u and $m \in S_u$.

[3] Link adaptations and turbo codes (implemented e.g. in UMTS release HSDPA) permit to approach the above *Shannon's* bound with the loss of a few dB in ξ.

C. Decentralized Congestion Control

We are interested in fair schemes for assigning transmission rates in elastic traffic, namely for a traffic that can accommodate bit-rate variations. We consider the case with no admission control, where an increase of the number of users in the network is just coped with via a reduction of the bit-rates of the users. Thus, we do not assume the bit-rates of users (or equivalently the ${\xi'}^u_m$ parameters) to be specified. Instead, we will look for fair schemes for assigning these rates (equivalently, parameters ${\xi'}^u_m$) under the constraints that the respective power control problem (3.4) is feasible. We now recall briefly, again taking the downlink as an example, the decentralized sufficient conditions for feasibility developed in [1], [2]. For user $m \in S_u$ define its *power-control load* with respect to BS u by $f^u_m = {\xi'}^u_m \sum_v \alpha_{uv} l^u_m / l^v_m$. Consider the following condition

$$\sum_{m \in S_u} f^u_m \le C\,, \tag{3.5}$$

for some constant $C > 0$.

Given some fair rate allocation policy (to be defined later), the following algorithms will be considered.

Congestion Control Protocol (CCP): Each BS periodically modifies the SINR's ξ^u_m (and thus, the bit-rates R^u_m) allocated to all mobiles in its cell according to some fair policy under the constraint of the condition (3.5).

It is shown in [2] that the application of CCP with the constant $C = 1 - \epsilon$ (for ϵ arbitrarily small) by all the BS's guarantees the global feasibility of the downlink power control problem (3.4) without power constraints. Moreover, it is shown there, that a similar condition (with some $C < 1$) guarantees the solution of the power allocation problem with maximal-power constraints (this condition, apart of the maximal BS powers, takes into account the external noises and common channel powers). The uplink is studied there as well. The proposed schemes are said to be *decentralized* in that each base station decides on the rates allocated based on its policy and on the location of the mobiles in its cell and the location of other base stations but not on the location of mobiles outside its own cell. Note that a particular bit-rate allocation decided by the CCP might consist in choosing only one user (e.g. with the best radio conditions) and assigning to him the whole channel capacity. Thus, a (opportunistic) time-division multiple-access is a special case of the system with power control.

IV. Instantaneous Fairness in Decentralized Congestion Control

In the previous section we have shown how the global feasibility of power control in a large network can be dealt with by imposing decentralized constraints on SINR threshold's, and hence instantaneous bit-rates. In the present section we discuss some classical *fair policies of instantaneous bit-rate assignment* under the decentralized constraints. Explicit formulas are available after some linearizing of the constraint.

A. Fairness and Optimality — Reminder

We now briefly remind the basic notions and facts concerning the fairness and optimality in resource allocation. We assume that we have N entities (think of mobile users in a given cell) which we index with $m = 1, \ldots, N$. The goal is to allocate resources (think of bit-rates) $\mathbf{R} = (R_1, \ldots, R_N)$ to these entities respecting some *constraint* in the form $\mathbf{R} \in \mathcal{R}$, where the set of *feasible allocations* $\mathcal{R}$ is some given subset of $\mathbb{R}^N$.

1) Optimality: An allocation $\mathbf{R} \in \mathcal{R}$ is called *(globally) optimal* if it maximizes $\sum_{m=1}^N R_m$. An allocation $\mathbf{R} \in \mathcal{R}$ is called *(strictly) Pareto optimal* if there is no solution $\mathbf{R}' \in \mathcal{R}$ dominating it.

2) Max-Min Fairness: [14] An allocation $\mathbf{R} \in \mathcal{R}$ is called *max-min fair* if for each $m \in \{1, \ldots, N\}$ increasing R_m must be at the expense of decreasing some already smaller R_n. If a max-min fair allocation exists on a set $\mathcal{R}$, then it is unique and strictly Pareto optimal.

3) Proportional Fairness: [15] An allocation $\mathbf{R} \in \mathcal{R}$ is called *proportionally fair* if for each other allocation $\mathbf{R}' \in \mathcal{R}$ we have $\sum_{m=1}^N (R'_m - R_m)/R_m \le 0$. Assume from now on that $\mathcal{R} \subset (\mathbb{R}^+)^N$, $\mathbb{R}^+ = \{r : r \ge 0\}$. If a proportionally fair allocation exists on $\mathcal{R}$, then it is unique and it is the solution of the following maximization problem $\max_{\mathbf{R} \in \mathcal{R}} \sum_{m=1}^N \log R_m$.

4) Generalized Fairness: [16] Consider the maximization problem $\max_{\mathbf{R} \in \mathcal{R}} \sum_{m=1}^N R_m^{1-\alpha}/(1-\alpha)$. Its solution is called the *generalized fair policy*. The following relations hold (see [16] for the proof).

Proposition 4.1: The general fair policy is globally optimal when $\alpha \to 0$, proportionally fair when $\alpha \to 1$, and max-min fair when $\alpha \to \infty$.

B. Linearizing the Constraints of the Decentralized Congestion Control

Consider the congestion control protocols CCP in Section III-C, applied by a given BS, say $u = 0$, located at $Y^0 = 0$. In what follows, we will omit in the notation the superscript $u = 0$. Suppose that at a given time there are N users served by this BS, and assume that all their characteristics, such as locations $\{X_m\}$, path losses l^v_m with respect to all BS's v (including $v = 0$) are fixed, as well as noises $N^u_m \equiv \bar{N}$ and orthogonality factors $\alpha_m \equiv \alpha$. The BS controls the load in its cell by assigning different (modified) SINR's ${\xi'}_m$. Note that the protocol CCP allows for configurations of modified SINR's $\xi' = \{\xi'_m\}$ $(m = 1, \ldots N)$ satisfying the condition of the form

$$\sum_{m=1}^N \xi'_m \gamma'_m \le C, \tag{4.1}$$

where γ'_m is a factor related to user's m reception/emission conditions (path losses, external noise). Note that the constraint (4.1) is linear in the modified SINR ξ'. However, we want to study rate allocation policies and thus we have to relate ξ' to the corresponding bit-rates $\mathbf{R} = \{R_i\}$. In order to simplify analysis, we will work under the following *linearizing assumption*:

(L) the bit-rate R is approximately some linear function of the SINR ξ; i.e. $R \approx \sigma\xi$ for some constant $\sigma > 0$.

Note by (3.1) that $\xi'_m \le \xi_m$ and thus under assumption (L) the following condition

$$\sum_{m=1}^N R_m \gamma_m \le 1, \tag{4.2}$$

with $\gamma_m = \sigma^{-1}\gamma'_m / C$, implies (4.1).

The assumption (L) is not very restrictive. Note that the relation (3.2) is linear. In this case we have $\sigma = W/(E_b/I_0)$. The assumption (L) is also asymptotically satisfied by Shannon's relation (3.3); indeed $\log_2(1+\xi) \approx \xi/\log 2$ for small SINR and in this case $\sigma = B^{-1}\log_2$. For such SINR's, also the inequality $\xi'_m \le \xi_m$ becomes tight, which means that the constraints (4.2) and (4.1) become equivalent.

Denote by $\mathcal{R}$ the set of all non-negative rate vectors $\mathbf{R} = \{R_m\}$ satisfying condition (4.2). We can identify now fair policies under this linear constraint.

Proposition 4.2: Consider the rate allocation problem under constraint (4.2). Policy $R_m = \gamma_m^{-1/\alpha}/\big(\sum_{j=1}^N \gamma_j^{1-1/\alpha}\big)$ *is the solution of the generalized fairness problem (see [17] for the proof). Consequently,*

$$R_m = \frac{1}{\sum_{j=1}^N \gamma_j} \text{ is the max-min fair allocation,} \quad (4.3)$$

$$R_m = \frac{1}{N\gamma_m} \text{ is the proportional fair allocation,} \quad (4.4)$$

$$R_m = \frac{1\{m \in J\}}{\sum_{j\in J}\gamma_j} \text{ is an optimal allocation,} \quad (4.5)$$

where J *is a non-empty subset of the set* $\{j : \gamma_j = \min_i \gamma_i\}$.

We will also consider a weighted modification of the optimal allocation (4.5)

$$R_m = \frac{1\{m \in J_w\}}{\sum_{j\in J_w}\gamma_j}\,, \quad (4.6)$$

where J_w is any non-empty subset of $\{j : w_j\gamma_j = \min_i w_i\gamma_i\}$ and $w = (w_1, \dots, w_M)$ are given weighting coefficients. Lets call (4.6) a *weight-based optimal allocation*. Note that it maximizes $\sum_m w_m R_m$ under constraint (4.2).

V. Model Dynamics

In this section we introduce the *traffic dynamics*. For this traffic, in the next section, we evaluate the average bit-rates (throughput) under different rate assignment policies. We suppose that there are *no handovers* (migration of users between cells). In such a situation, our decentralized congestion control principle allows one to consider users in one typical cell and study their throughput, which depends on the geometry of the whole network.

We consider a spatial Markov queuing model for elastic traffic, called WET model, that is described and analyzed in the Appendix. It assumes Markovian arrivals, departures and mobility of customers, which transmit some given data-volumes. We enrich this model with independent fluctuations of the channel conditions of each user. Thus we have the following three elements of the model dynamics: *fading*, *arrivals/departures*, *mobility of users*, which we describe now in details.

A. Mean Path-Loss and Fast Fading

From now on we assume that the path gain $1/l_m^0$ of each user m with respect to its own BS $u = 0$ located at $Y^0 = 0$, varies in time around some *mean path gain*, which depends only on the distance between this user and its BS; i.e., $l_m^0 = \bar{l}_m^0/F_m(t)$, where $\bar{l}_m^0 = L(|Y^0 - X_m^0|)$. The function $L(r)$ will be called the *(mean) path loss* (averaged over the fading effects; in fact it is the inverse of the mean path gain) and the process $F_m(t)$ is called *fading*. We will always assume that the fading processes of different users are independent, identically distributed, ergodic processes with mean 1.

Finally, we make a general assumption that the fading does not play any essential role for the interference. One can think that many BS's contribute in an additive way to the interference (cf. (3.4)). Thus, individual fading conditions on the links between these BS's and a given user are averaged out (see [18, §4.3.1] for more detailed arguments). This means that we take $l_m^v = L(|Y^v - X_m^v|)$ for $v \ne 0$.

B. Mobility

Denote by $\mathbb{D} \subset \mathbb{R}^2$ the cell served by the given BS $u = 0$. We will model the internal mobility of users in $\mathbb{D}$ by some Markov process. Specifically, assume that users move independently of each other in $\mathbb{D}$. The sojourn duration of a given user at location $x \in \mathbb{D}$ is exponentially distributed with parameter $\lambda'(x)$. Any user finishing its sojourn at location x, is routed to a new location dy according to some probability kernel $p'(x, \mathrm{d}y)$, where $p'(x, \mathbb{D}) = 1$.

The above description corresponds to a Markov process on $\mathbb{D}$ with the following generator (of the individual user mobility): $\lambda(x, \mathrm{d}y) = \lambda'(x)p'(x, \mathrm{d}y)$. We will always assume that this Markov process is ergodic and will denote its invariant measure by $\rho_{mob}(\cdot)$; it satisfies the following traffic equations: $\rho_{mob}(\mathbb{D}) = 1$ and

$$\int_A \lambda(x, \mathbb{D})\,\rho_{mob}(\mathrm{d}x) = \int_{\mathbb{D}} \lambda(x, A)\,\rho_{mob}(\mathrm{d}x)\,, A \subset \mathbb{D}\,. \quad (5.1)$$

In fact, the above Markovian structure is not essential when studying the network performance in two limiting regimes: small mobility and infinite mobility, which we will be interested in at Section VI. Results for the latter regime depend only on the stationary distribution $\rho_{mob}(\cdot)$.

C. Arrivals and Departures

We will model the process of call arrivals to and departures from $\mathbb{D}$ as a spatial birth-and-death (SBD) process: for a given subset $A \subset \mathbb{D}$, interarrival times to A are independent of everything, exponential random variables with mean $1/\lambda(A)$, where $\lambda(\cdot)$ is some given intensity measure of arrivals to $\mathbb{D}$ per unit of time. This allows the modeling of spatial hot spots. In homogeneous traffic conditions, we can take $\lambda(\mathrm{d}x) = \lambda\mathrm{d}x$, where λ is the mean number of arrivals per unit of area and per unit of time. We assume that each arrival brings to the system some job-volume (amount of bits that has to be sent or received), which is modeled by an exponential variable of parameter τ, independent of everything else. Users are served by the BS according to some bit-rate assignment policy.

The above three types of dynamics (fading, mobility and arrivals/departures) often coexist on different time scales. A right separation of their time scales in modeling allows for explicit network performance evaluation, as we show in the next section.

VI. Performance Evaluation

In this section we will analyze the mean number of users, the mean throughput and the mean delay obtained by users in one typical cell of some (possibly large) network. For this we assume the WET model of arrivals departures and user-mobility, enriched by independent fluctuations of the channel conditions of each user. These elements of the model were described in the previous section. We begin with a principle that allows for an explicit evaluation of the performance of the network with a multi-scale dynamics.

Separation of the Time Scales: Consider two types $\mathcal{X}, \mathcal{Y}$ of the traffic dynamics (think of $\mathcal{X}, \mathcal{Y} \in \{$ fading (F), mobility (M), arrivals/departures (AD)$\}$). Let one of them, say $\mathcal{X}$ be much faster than the other (denote it $\mathcal{X} \prec \mathcal{Y}$). Precisely, suppose that it is reasonable to assume that $\mathcal{X}$ reaches its steady state regime in a time that is much shorter than the typical period "between two successive changes of the configuration of $\mathcal{Y}$". In such a situation one may consider the following *two-step performance evaluation*. First, freeze $\mathcal{Y}$ and for this given configuration study the steady state performance $\mathcal{X}$. Then, let $\mathcal{Y}$ operate and consider the steady-state characteristics of $\mathcal{X}$ as instantaneous performance of $\mathcal{Y}$. Finally, study the steady state performance of $\mathcal{Y}$. These higher-level performance characteristics may be taken as an approximation of the performance of the real multi-scale dynamics.

The above idea was used e.g. in [9] to evaluate the performance of the channel-aware rate scheduling under a dynamical user configuration. Following it, in what follows we will assume that the channel fluctuations (fading) are much faster than arrivals/departures of users (F$\prec$AD). Moreover, we will consider the following two extreme scenarios concerning this latter variability: *no mobility*; i.e., only arrivals and departures change the user configuration, and *infinite mobility*, meaning that we separate the time scale of the mobility and the arrivals/departures (M$\prec$AD).

In what follows we will be interested in two particular instantaneous rate assignment policies: a weight-based optimal policy (4.6) and proportional fair policy (4.4). The first was shown (see e.g. [10]) to be a good model for opportunistic schedulers, like in HSDPA (or HDR). The second one can be used to analyze non-opportunistic schedulers, like in DSCH of the UMTS release 99.

A. Performance of the Channel-Aware Scheduling

We assume that the channel fading is faster than arrivals/departures (F$\prec$AD). Moreover, in this section we suppose that the BS rate scheduler can follow the fluctuation of the channel fading and we evaluate its performance for a given fixed configuration of users (number and positions).

Consider our linear constraint (4.2) and remember that $\gamma_m = \sigma^{-1}\gamma'_m/C$. Note that in the congestion control protocol CCP considered in Section III-C, the coefficient γ'_m is some multiple of the path-loss of the user m with respect to its BS $u = 0$. By the assumption on the path-loss and fading made in Section V-A we have

$$\gamma'_m = \bar{\gamma}'_m / F_m(t)\,, \tag{6.1}$$

where $F_m(t)$ is the fading process of the user m with respect to its serving BS, and $\bar{\gamma}'_m$ is given by the same formula as γ'_m (the respective power control load formula of Section III-C) with l^v_m replaced by $\bar{l}^v_m = L(|Y^v - X^v_m|)$ for all v.

Note that by (4.2) and (6.1), the *feasible rate* at time t of user m (if it were the only user) is equal $r_m(t) = F(t)/\bar{\gamma}_m$ where $\bar{\gamma}_m = \sigma^{-1}\bar{\gamma}'_m/C$. We have assumed that $F(t)$ is ergodic and $\mathbf{E}[F(t)] = 1$, so the *averaged feasible bit-rate* of user m equals to $\bar{r}_m = 1/\bar{\gamma}_m$.

1) An Opportunistic Weight-Based Optimal Policy: Following [9], consider now a particular weight-based policy (4.6) with $w_m = 1/\bar{r}_m = \bar{\gamma}_m$, which offers to user m the instantaneous rate

$$R_m(t) = \frac{F_m(t)}{\bar{\gamma}_m} 1\Big(F_m(t) = \max_{j=1,\dots,N} F_m(t)\Big)\,; \tag{6.2}$$

(we consider a suitable tie-breaking rule if the above channel maximization problem does not have a unique solution). By the ergodicity of the fading processes, independence and the symmetry, one can show (cf. [9]) that this weight-based policy gives to any user m of the fixed configuration of N users served by the BS the *throughput*

$$\bar{R}_m = \mathbf{E}[R_m(t)] = \frac{G(N)}{N\bar{\gamma}_m}\,, \tag{6.3}$$

where $G(N) = \mathbf{E}[\max(F_1, \dots, F_N)]$ and F_i are independent, identically distributed copies of the fading variable.

2) Non-opportunistic Proportional Fair Policy: Consider now the situation when the BS scheduler applies the instantaneous proportional fair policy (4.4). By (6.1) we have the instantaneous rate offered to user m equal to

$$R_m(t) = \frac{F_m(t)}{N\bar{\gamma}_m} \tag{6.4}$$

and thus its throughput is equal to

$$\bar{R}_m = \mathbf{E}[R_m(t)] = \frac{1}{N\bar{\gamma}_m}\,. \tag{6.5}$$

Remark: Comparing (6.3) to (6.5) we notice the throughput gain of the factor $G(N)$ for any fixed configuration of N users (note that $G(n) > \mathbf{E}[F_1(t)] = 1$) of the opportunistic weight-based policy (6.2) with respect to the non-opportunistic proportional fair policy (6.4). Moreover, it can be directly shown that the weight-based policy (6.2) maximizes $\sum_{m=1}^N \bar{\gamma}_m \log \bar{R}_m$ given the instantaneous rate constraint (4.2) (see [17]) for a direct proof). Thus we say that it realizes the *weighted proportional fairness of the throughput* (a long-term fairness) [4] given the instantaneous rate constraint. Note that the non-opportunistic proportional fair policy (6.4) realizes the weighted proportional fairness of the *instantaneous rates* (instantaneous fairness), and consequently it is less performant in the long-term run. In order to avoid possible confusions we will call the policy (6.2) *the opportunistic proportional fair policy*.

B. Performance of a Variable, Motionless Population

In this section we assume only arrivals and departures of users in the domain $\mathbb{D}$, as described in Section V-C, who do not move during their service. We assume that these events happen on a time scale that is much slower than the fading process (F$\prec$AD). By the principle of the separation of the time scales, we will consider the channel throughput characteristics (6.3), (6.5), calculated for fixed user configurations, as instantaneous bit-rates offered by the BS on the time scale of the arrivals and departures.

Specifically, we consider the WET process with generator given by (A.1), where $\lambda(x, \mathbb{D}) = 0$ and the service rate $r_x(\nu)$ is given by (A.2) where $\nu(\cdot)$ is a point measure describing the locations of users. Following considerations of Section V-A we

[4]which justifies the name *proportional fair* given to the HSDPA (HDR) algorithm

take $\bar{\gamma}(x) = \sigma^{-1}\bar{\gamma}'(x)/C$, where $\bar{\gamma}'(x)$ is the respective power control load of user m located at x (see the respective power control load formula of Section III-C) taken with respect to the mean path-loss $l_m^v = L(|Y^v - X_m^v|)$ for all v. The traffic intensity measure $\rho(\cdot)$ is given by (A.3).

Depending on the service policy we take $h(n) = n/G(n)$ in the case of opportunistic proportional fair one (6.3) and $h(n) = n$ for the non-opportunistic proportional fair policy (6.5). Denote

$$\mathcal{H}(s) = \frac{\sum_{n=0}^{\infty} s^n (n+1) \prod_{i=1}^{n+1} G(i)}{\sum_{n=0}^{\infty} s^n \prod_{i=1}^{n} G(i)} . \qquad (6.6)$$

The following result follows from Propositions A.1 and A.3.

Proposition 6.1:
- *The WET model with non-opportunistic proportional fair policy (6.5) is stable for $\rho(D) < 1$. The mean number of users in A in steady state of the system is equal to $X(A) = \rho(A)/(1 - \rho(\mathbb{D}))$. The expected delay is equal to $V(A) = \rho(A)/(\lambda(A)(1 - \rho(\mathbb{D})))$ and the mean throughput $T(A) = \lambda(A)(1 - \rho(\mathbb{D}))/(\tau\rho(A))$.*
- *The WET model with opportunistic proportional fair policy (6.3) is stable for $\rho(D) < \lim_{n\to\infty} G(n)$. Its performance characteristics $X(A), V(A), T(A)$ are given by Proposition A.3 with $\mathcal{H}(\cdot)$ given by (6.6).*

C. Performance of a Variable Population with Infinite Mobility

In this section we assume that the mobility of users is much faster than the time scale of the arrivals and departures of users (M≺AD). Then, three different scenarios are possible:

- F,M≺AD and the scheduler applies the non-opportunistic proportional fair policy,
- F≺M≺AD and the scheduler applies the weight-based opportunistic policy at the level of fading only,
- F≈M≺AD and the scheduler applies the weight-based opportunistic policy jointly with respect to fading and mobility.

By the principle of the separation of the time scales, we will consider the channel throughput characteristics (6.3), (6.5) calculated for fixed *number* of users as instantaneous bit-rates offered by the BS on the time scale of the arrivals and departures.

Specifically, we consider the WET process with generator given by (A.1). As in Section A-3, we assume a constant arrival intensity $\lambda(\mathrm{d}x) = \lambda(\mathbb{D})\,\mathrm{d}x/|\mathbb{D}|$ ($|\mathbb{D}|$ is the surface area of $\mathbb{D}$) and the service rates averaged over the stationary distribution ρ_{mob} of user location; i.e, service rates $r(\nu)$ given by (A.5). We take $\bar{\gamma}(x)$ as in Section VI-B. The traffic intensity ρ is given by (A.6).

Depending on the scenario considered above and the service policy, we take

- F,M≺AD: $h(n) = n$,
- F≺M≺AD: $h(n) = n/G(n)$,
- F≈M≺AD: $h(n) = n/G'(n)$, where $G'(n) = \mathbf{E}[\max(F_1/\bar{\gamma}(X_1), \ldots, F_n/\bar{\gamma}(X_n)]$ with F_i, X_i being independent, identically distributed copies of, respectively, the fading variable, and the stationary user location (i.e, X_i is distributed according to ρ_{mob}).

The following result follows from Propositions A.2 and A.3.

Proposition 6.2:
- *The WET model with F,M≺AD is stable for $\rho < 1$. The mean number of users in steady state of the system is equal to $X = \rho/(1 - \rho)$. The expected delay is equal to $V = \rho/(\lambda(\mathbb{D})(1 - \rho))$ and the mean throughput $T = \lambda(\mathbb{D})(1 - \rho)/(\tau\rho)$.*
- *The WET model with F≺M≺AD is stable for $\rho < \lim_{n\to\infty} G(n)$. Its performance characteristics X, V, T are given be Proposition A.3 with $\mathcal{H}(s)$ given by (6.6).*
- *The WET model with F≈M≺AD is stable for $\rho < \lim_{n\to\infty} G'(n)$. Its performance characteristics X, V, T are given by Proposition A.3 with $\mathcal{H}(s)$ given by (6.6) and G replaced by G'.*

Remark: Comparing the results given in Propositions 6.1 and 6.2, by Proposition A.4 and Corollary A.5 one finds that the opportunistic proportional fair policy performs better than the non-opportunistic one in both motionless and infinite mobility scenario. Moreover, mobility increases the performance of both policies. The numerical examples are given in the next section.

The multi-class extension of the model can be studied in a similar manner (see [17]). The performance can be evaluated explicitly and it can be shown that the policy realizing the proportional fairness between the classes of users gives larger mean individual user delays than the policy (4.4), which assures the same type of fairness between users.

One can also study a model with infinite but local mobility see Remark after Corollary A.5, which might be a more pertinent assumption than the infinite mobility over the entire cell.

Finally, one can consider the streaming and elastic traffic sharing the same channel. Giving priority to the streaming, one can evaluate the performance of the fair elastic-rate assignment policies, discussed previously, applied to the remaining part of the channel. This approach consists in replacing C in (4.1) by $C - \sum_n \xi'_n \gamma'_n$, where the summation is taken over the streaming-traffic users and ξ'_n are SINR's corresponding to their predefined bit-rates, see [17] for the details.

VII. Numerical Results

In this section we will give a few numerical examples for a hexagonal network.

In order to evaluate numerically the performance of the network we have to specify the geometry of BS's, the user mobility, the path-loss, the fading, as well as the coding and the service policy.

We consider *an infinite hexagonal network of BS's*. Each BS serves users in its *cell* defined as the set of locations in the plane which are closer to that BS than to any other BS. We model path-loss on distance r by $L(r) = (Kr)^\eta$, and take in particular $\eta = 3.38$, $K = 8667$.

The above assumptions on the network architecture and the path-loss allow to approximate the function $\bar{\gamma}'(x)$ as a linear operator of the following function

$$f(x) = \zeta(\eta-1)L(|x|)\Big(\frac{1}{L(\Delta - |x|)} + \frac{1}{L(\Delta + |x|)} + \frac{4}{L\Big(\sqrt{\Delta^2 + |x|^2}\Big)}\Big)$$

for $|x| \le R$, where $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ is the Riemann zeta function and Δ is the distance between two adjacent BS's in the hexagonal network and R is the radius of the disc with area equal to that of the cell (see [19]). Note that $\Delta^2 = 2\pi R^2/\sqrt{3}$

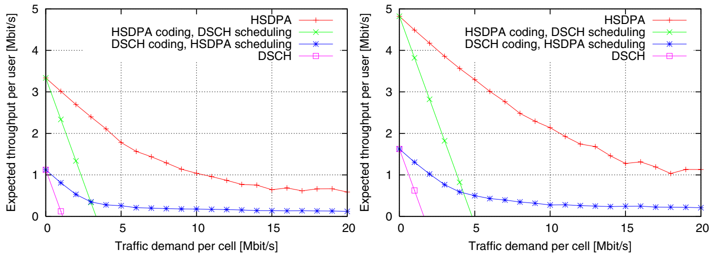

Fig. 1. Mean throughput for different policies in motionless scenario (left) and infinite mobility (right)

and $\lambda_{BS} = 1/(\pi R^2)$. The form of this linear operator ($\bar{\gamma}'$ of f) depends on the protocol. Specifically for the downlink CCP without maximal power constraints $C = 1 - \epsilon$, $\bar{\gamma}'(x) = \alpha + f(x)$, Recall that $\bar{\gamma}(x) = \sigma^{-1}\bar{\gamma}'(x)/C$ and σ depends on the coding.

We consider two different codings, for which the respective values of the factor σ (cf. linearizing assumption (L) in Section IV-B: $\sigma = 1.21$Mbit/s, which corresponds to DSCH coding and $\sigma = 3.61$Mbit/s, which corresponds to HSDPA coding. We take the orthogonality factor $\alpha = 0.4$ that corresponds to the UMTS.

We assume the *Rayleigh fading*; i.e., the distribution of $F(t)$ for any given t is exponential with mean 1 (see e.g. [18, p. 50 and 501]). In this case $G(n) = \sum_{j=1}^{n} 1/j$.

We assume the uniform stationary distribution $\rho_{mob}(\mathrm{d}x) = \mathrm{d}x/(\pi R^2)$ of user mobility.

Figure 1 shows the mean throughput for different scheduling policies in the motionless and infinite mobility scenario (F,M≺AD for DSCH and F≺M≺AD for HSDPA). Note that the non-opportunistic policy gives the throughput that is a linear function of the demanded traffic in the bounded stability region, while the opportunistic one is stable on the whole line, and its gain increases with the traffic. The HSDPA coding shows the gain of the factor 3. Moreover, the infinite mobility increases the throughput of about 1.5 with respect to the static scenario.

Concluding Remarks

We have given analytic expressions for the mean number of users, the mean throughput and the mean delay in power controlled cellular networks with elastic traffic. These formulas take into account in a simple but not simplistic way all important elements of the network performance: the network geometry, coding, service rate policy, user mobility, and allow one to compare in a systematic way the performance gain obtained at the different layers (better coding, channel-aware opportunistic service policy, fast user mobility).

An important fact is that the fair and feasible transmission rates assigned by the considered service policies, are deduced from schemes, which ensure globally in the network existence of the respective powers that should be allocated in order to obtain these transmission rates. These schemes, based on an exact representation of the geometry of channels, can be implemented in such a way that each base station only has to consider the load brought by its own users.

Different model assumptions used in this paper have already been validated separately. In particular our scalable load control schemes taking into account the geometry of interferences have been validated in [2] by comparison with simulations, in the case of the feasibility of the power allocation (in the context of a streaming traffic). *Orange* already implemented in its dimensioning tools our analytical expressions for both the streaming and the elastic traffic.

Acknowledgment

This paper originates from a long collaboration of the authors with François Baccelli, within which the previous results on the scalable admission/congestion control results were obtained. The authors are grateful to FB for many helpful discussions. Special thanks are also to Omar Soufit for numerical evaluation of the scalable admission/congestion schemes.

Appendix: Mathematical background

In this section we develop mathematical tools for the spatio-temporal analysis of the wireless networks with elastic traffic. In particular we introduce and study a spatial *Markov Wireless Elastic Traffic* (WET) model. It can be seen as an extension of the Spatial Queueing System described in [4, Chapter 10], which allows for departure rates depending, not only on the current, but also on the next location of a given user (besides current locations of other users).

1) WET Model: Very much as in [4], we will consider a system in which units (users) take place in a complete, separable metric space $\mathbb{D}$, where they are processed. We will represent the state of the system by a finite *counting measure* ν on $\mathbb{D}$. A random configuration N of units at a given time, will be modeled by a *point process* that is a measurable mapping from some given probability space to the state space $\mathbb{M}$ of all finite counting measures on $\mathbb{D}$ with an appropriate σ-algebra $\mathcal{M}$.

Remind that for or a given nonnegative function $\Psi : \mathbb{M} \to \mathbb{R}_+$ ($\mathbb{R}_+$ denotes the set of non-negative real numbers) and a mean measure λ on $\mathbb{D}$, the *Gibbs distribution* on $\mathbb{M}$, with *energy function* Ψ and Poisson *weight process* N of mean measure ρ, is the distribution Π on $\mathbb{M}$ defined by $\Pi(\Gamma) = Z^{-1}\mathbf{E}[1(N \in \Gamma)\Psi(N)]$, where $Z = \mathbf{E}[\Psi(N)]$ is the normalizing constant called also *partition function* or *statistical sum*.

We will need a "virtual" state $o \notin \mathbb{D}$ which can be seen as a location outside the space $\mathbb{D}$, and which represents the initial location of particles arriving to or leaving the system. Denote $\overline{\mathbb{D}} = \mathbb{D} \cup \{o\}$.

Define the following *displacement operator* T on the space $\mathbb{M}$ for $\nu \in \mathbb{M}, x, y \in \mathbb{D}$: $T_{oy}\nu = \nu + \varepsilon_y$, $T_{xo}\nu = \nu - \varepsilon_x$ defined only if $\nu(\{x\}) \geq 1$, and $T_{xy}\nu = \nu - \varepsilon_x + \varepsilon_y$ defined only if $\nu(\{x\}) \geq 1$.

The transition $\nu \to T_{ox}\nu$ will be called the *arrival* of a user at x; the transition $\nu \to T_{xo}\nu$ is the *departure* of a user from x, which is well defined provided $\nu(x) > 0$; the transition $\nu \to T_{xy}\nu$, $x, y \in \mathbb{D}$ is the displacement of a user from x to y.

Define the WET model as a spatial Markov queueing process with the following infinitesimal generator : for $\nu \in \mathbb{M}$,

$\Gamma \in \mathcal{M}$ let

$$q(\nu,\Gamma) = \int_{\mathbb{D}} 1(T_{oy}\nu \in \Gamma)\,\lambda(\mathrm{d}y) + \int_{\mathbb{D}} \tau r_x(\nu) 1(T_{xo}\nu \in \Gamma)\,\nu(\mathrm{d}x) + \int_{\mathbb{D}}\int_{\mathbb{D}} 1(T_{xy}\nu \in \Gamma)\,\lambda(x,\mathrm{d}y)\,\nu(\mathrm{d}x)\,, \quad \text{(A.1)}$$

where $\lambda(A)$ is the *arrival rate* (from o) to $A \subset \mathbb{D}$; (we assume $0 < \lambda(\mathbb{D}) < \infty$), τ is the parameter of the exponential *job-volume* (amount of data that are to be sent) brought to the system by each user, $\lambda(x, A)$ ($x \in \mathbb{D}$, $A \subset \mathbb{D}$) is a Markov *mobility kernel* on the state space $\mathbb{D}$ that describes the mobility of a single user; (we implicitly assume that $0 \le \lambda(x,\mathbb{D}) < \infty$ and $\lambda(x,\{x\}) = 0$ for all $x \in \mathbb{D}$), $r_x(\nu)$ ($x \in \mathbb{D}$, $\nu \in \mathbb{M}$, $\nu(\{x\}) > 0$) is the *transmission rate* of a user located at x in the configuration ν of all users; (we implicitly assume that $0 \le r_x(\nu) < \infty$). Note that, by the definition, q is conservative and stable.

Our WET model for wireless networks serving elastic traffic is different form the Whittle model proposed for such networks in [3, § 2]. In fact, note that the WET model is not a Whittle network since the departure rate from x depends on ν *and* on the next position of the user (see Section II for more discussion). If the WET model is more realistic, its inconvenience is that we cannot explicitly evaluate its performance. For this reason, following the idea of [5], we consider here only the following two extreme cases: *no mobility* and *infinite mobility*. In the former case the WET is equivalent to a spatial version of the Whittle model of [3, § 2], while in the latter case we proceed by a *separation of the time scales* of mobility and arrivals/departures. In both cases we will assume a product form of the service rates

$$r_x(\nu) = \frac{1}{h(\nu(\mathbb{D}))\bar{\gamma}(x)}\,, \quad \text{(A.2)}$$

where $h(n)$ is some nonnegative function of the total number of users $\nu(\mathbb{D})$ in service, while $\bar{\gamma}(x)$ is a nonnegative function of the user location $x \in \mathbb{D}$. This form allows for explicit evaluation of the WET model in both extreme cases. Moreover, by the celebrated insensitivity property, this performance depends only on the mean job-load $1/\tau$ of each user and not on its distribution.

2) No Mobility Case: Assume now that there is no mobility of users in $\mathbb{D}$; i.e, $\lambda(x,\mathbb{D}) = 0$, and all users are served at their arrival location. In this case the WET model becomes a *spatial birth-and-death model that is equivalent to a spatial processor sharing queue*. Define the *traffic intensity measure*; i.e., traffic intensity per region $A \subset \mathbb{D}$, by

$$\rho(A) = \frac{1}{\tau}\int_A \bar{\gamma}(x)\,\lambda(\mathrm{d}x)\,. \quad \text{(A.3)}$$

Note that the location dependent factor $1/\bar{\gamma}(x)$ of the service rate is taken into account in ρ. Note also that the remaining, population-dependent factor $1/h(\nu(\mathbb{D}))$ of the service rate is balanced by the following function

$$\Psi(\nu) = \prod_{i=1}^{\nu(\mathbb{D})} h(i) \quad \text{(A.4)}$$

that depends only on the total number of users $\nu(\mathbb{D})$. With a slight abuse of the notation, we will sometimes write $\Psi(n)$ for $\Psi(\nu)$ with $\nu(\mathbb{D}) = n$. Denote by Z the normalizing constant

$$Z = e^{-\rho(\mathbb{D})}\sum_{n=0}^{\infty}\frac{\rho(\mathbb{D})^n}{n!}\prod_{i=1}^{n} h(n)\,.$$

Proposition A.1: The WET model without mobility is stable under the condition $\rho(\mathbb{D}) \le \lim_{n\to\infty} n/h(n)$ and the stationary distribution of the configuration of users in $\mathbb{D}$ is the Gibbs distribution with local energy $\Psi(\cdot)$ based on the Poisson weight process with mean measure $\rho(\cdot)$.

Proof: The result follows form the fact that in this case the WET model is a spatial birth-and-death model equivalent to a spatial processor sharing queue; see e.g. [6], [7] for stability and [4] for the invariant measure. ■

3) Infinite Mobility: Assume a constant arrival intensity $\lambda(\mathrm{d}x) = \lambda(\mathbb{D})\,\mathrm{d}x/|\mathbb{D}|$. Moreover, suppose that the mobility of users is so fast that it we can reasonably assume that during the periods of time when the number of users $\nu(\mathbb{D})$ is constant, each user receives a service rate that is averaged over its mobility; i.e., assume

$$r(\nu) = \int_{\mathbb{D}} r_x(\nu)\,\rho_{mob}(\mathrm{d}x) = \frac{\int_{\mathbb{D}}\frac{1}{\bar{\gamma}(x)}\,\rho_{mob}(\mathrm{d}x)}{h(\nu(\mathbb{D}))}\,, \quad \text{(A.5)}$$

where ρ_{mob} is the stationary distribution of the location of individual user in $\mathbb{D}$ given by (5.1) with $\rho_{mob}(\mathbb{D}) = 1$. Note that in this case the WET model is a *discrete birth-and-death model equivalent to a classical processor sharing queue*. Define its *traffic intensity* by

$$\rho = \frac{\lambda(\mathbb{D})}{\tau\int_A \frac{1}{\bar{\gamma}(x)}\,\rho_{mob}(\mathrm{d}x)}\,. \quad \text{(A.6)}$$

The population-dependent factor $1/h(n)$ of the service rate is balanced by the function $\Psi(n)$ given by (A.4). Denote by Z_∞ the normalizing constant

$$Z_\infty = e^{-\rho}\sum_{n=0}^{\infty}\frac{\rho}{n!}\prod_{i=1}^{n} h(i)\,.$$

Proposition A.2: The WET model is stable under the condition $\rho \le \lim_{n\to\infty} n/h(n)$ and the stationary distribution of the number of users in $\mathbb{D}$ has the density $\Psi(n)/Z_\infty$, with respect to the Poisson random variable with parameter ρ.

Proof: The result follows form the fact that in this case the WET model is equivalent to a discrete processor sharing queue; see e.g. [6] for stability and [4] for the invariant measure. ■

4) Performance Evaluation: Denote by $X(A)$ the expected number of users in set $A \subset \mathbb{D}$ in steady state of the WET model without mobility and denote by X the expected number of users in $\mathbb{D}$ in the the steady state of the WET model with infinite mobility. Similarly, let $V(A)$ and V be the *expected delay* of jobs in, respectively, set A in motionless model, and in $\mathbb{D}$ in the model with infinite mobility. Recall, by the Little's theorem (cf. [20]) that $V(A) = X(A)/\lambda(A)$ and $V = X/\lambda(\mathbb{D})$. Moreover, denote by $T(A)$ and T the *mean throughput (in the set A)* in the respective models. Specifically,

$$T(A) = \frac{\mathbf{E}_\Pi\left[\int_{\mathbb{D}} r_x(N(\cdot))\,N(\mathrm{d}x)\right]}{X(A)}\,,\qquad T = \frac{\mathbf{E}_\Pi[r(N)N]}{X}\,,$$

where $\mathbf{E}_\Pi$ is the stationary distribution of the point process $N(\cdot)$ of users in the static model and of the number of users

N in the infinite-mobility model. Define a function $\mathcal{H}(s)$ for $s > 0$

$$\mathcal{H}(s) = \frac{\mathbf{E}[\Psi(N_s + 1)]}{\mathbf{E}[\Psi(N_s)]}, \tag{A.7}$$

where N_s is a Poisson random variable with parameter s.

Proposition A.3: Assume that both static and infinitely mobile WET models are stable. Then we have for the respective models (see [17] for the detailed proofs)

$$X(A) = \rho(A)\mathcal{H}(\rho(\mathbb{D}))\,, \quad X = \rho\mathcal{H}(\rho)$$
$$V(A) = \frac{\rho(A)}{\lambda(A)}\mathcal{H}(\rho(\mathbb{D}))\,, \quad V = \frac{\rho}{\lambda(\mathbb{D})}\mathcal{H}(\rho)\,,$$
$$T(A) = \frac{\lambda(A)}{\tau\rho(A)\mathcal{H}(\rho(\mathbb{D}))}\,, \quad T = \frac{\lambda(\mathbb{D})}{\tau\rho\mathcal{H}(\rho)}\,.$$

Note that we have a simple relation between delay and throughput: $T(A) = 1/(V(A)\tau)$.

5) Comparison: Consider two stable WET models without mobility and/or with infinite mobility. Denote by $h_i(\cdot), \rho_i = \rho_i(\mathbb{D})$ the service rate function and the (total) intensity for the two models $i = 1, 2$. Assume the same arrival and job-volume intensities $\lambda(\cdot), \tau$ in both models. Denote by N_i the (random) number of users in steady sate of the respective models, and by X_i, V_i, T_i, respectively, the expected number of users, mean delay and the mean throughput in both models $i = 1, 2$. The following result allows us to compare the performance of the above models.

Proposition A.4: For the WET models described above, if $h_1(n)\rho_1 \le h_2(n)\rho_2$ for all $n \ge 1$ then N_1 is stochastically smaller than N_2. Consequently, $X_1 \le X_2$, $V_1 \le V_2$, $T_1 \ge T_2$.

Proof: We show that for any nondecreasing function $\phi(\cdot)$, $\mathbf{E}_{\rho_1}[\phi(N_1)\psi_1(N_1)]/\mathbf{E}_{\rho_1}[\psi_1(N_1)] \le \mathbf{E}_{\rho_2}[\phi(N_2)\psi_2(N_2)]/\mathbf{E}_{\rho_2}[\psi_2(N_2)]$, where ψ is given by (A.4) and $\mathbf{E}_{\rho_i}$ is the expectation with respect to the Poisson distribution $\mathbf{P}_{\rho_i}$. Note that $(\mathrm{d}\mathbf{P}_{\rho_2}/\mathrm{d}\mathbf{P}_{\rho_1})(N) = e^{\rho_1-\rho_2}(\rho_2/\rho_1)^N$. Consequently, the above inequality is equivalent to

$$\frac{\mathbf{E}_{\rho_1}[\phi(N_1)\psi_1(N_1)]}{\mathbf{E}_{\rho_1}[\psi_1(N_1)]}\frac{\mathbf{E}_{\rho_1}[\prod_{i=1}^{N_2}\frac{\rho_2 h_2(i)}{\rho_1 h_1(i)}\psi_1(N_2)]}{\mathbf{E}_{\rho_1}[\psi_1(N_2)]} \le \frac{\mathbf{E}_{\rho_1}[\phi(N_2)\prod_{i=1}^{N_2}\frac{\rho_2 h_2(i)}{\rho_1 h_1(i)}\psi_1(N_2)]}{\mathbf{E}_{\rho_1}[\psi_1(N_2)]}\,,$$

which holds by the association principle; see e.g. [21, § 3.10]. ■

Now, in order to compare the performance of the model with null and infinite mobility, assume that in the static model the arrivals take locations in $\mathbb{D}$ according to the stationary distribution of the locations of individual users in the dynamic model; i.e., $\lambda(\mathrm{d}x)/\lambda(\mathbb{D}) = \rho_{mob}(\mathrm{d}x)$.

Corollary A.5: Given the total arrival rate, job-volume and the same mean repartition of the population in $\mathbb{D}$, the expected number of users in the system with infinity mobility is smaller than in the model without mobility.

Proof: We have $\rho(\mathbb{D}) = \lambda(\mathbb{D})\int_{\mathbb{D}}\bar{\gamma}(x)\,\rho_{mob}(\mathrm{d}x)/\tau$ and by the Jensen's inequality that $\rho \le \rho(\mathbb{D})$. The result follows from Proposition A.4 ■

Note that this results was established in a more general context in [5]. Moreover, it is shown there that the monotonicity holds with respect to the speed of mobility on the whole half-line $[0, \infty]$ for the model considered there.

Remark: One can consider the following extension of the WET model with *local mobility*. Suppose that users arrive to the system as before and, depending on its arrival location x, each user moves according to some local dynamics that has a stationary regime $\rho^x_{mob}(\cdot)$. As before, one can consider two cases: no mobility and infinite local mobility. The former case is equivalent to some motionless WET model studied in Section A-2, while the latter one consists in a modification of the same motionless model, where the service rates $r_x(\nu)$ are averaged over the local mobility; i.e., taken to be $r_x(\nu) = \int_{\mathbb{D}} 1/\bar{\gamma}(y)\,\rho^x_{mob}(\mathrm{d}y)/h(\nu(\mathbb{D}))$ As previously, one can show that such a local mobility enhances the system performance.

REFERENCES

[1] F. Baccelli, B. Błaszczyszyn, and F. Tournois, "Downlink capacity and admission/congestion control in CDMA networks," in *Proc. of IEEE INFOCOM*, San Francisco, 2003.

[2] F. Baccelli, B. Błaszczyszyn, and M. Karray, "Up and downlink admission/congestion control and maximal load in large homogeneous CDMA networks," *MONET*, vol. 9, no. 6, pp. 605–617, December 2004.

[3] T. Bonald and A. Proutière, "A queueing analysis of data networks," in *Queueing Networks: A Fundamental Approach*, Boucherie and Van Dijk, Eds., 2006.

[4] R. Serfozo, *Introduction to stochastic networks*. New York: Springer, 1999.

[5] T. Bonald, S. Borst, and A. Proutiére, "How mobility impacts the flow level performance of wireless data systems," in *In Proc. of IEEE Infocom*, 2004.

[6] J. Cohen, *On Regenerative Processes in Queueing Theory*, ser. Lecture Notes in Economics and Mathematical Systems. Springer Verlag, 1976, vol. 121.

[7] J.-M. Kelif, I. Koukoutsidis, and E. Altman, "Fair rate sharing models in CDMA link with multiple classes of elastic traffic," INRIA, Tech. Rep. RR 5596, 2005.

[8] C. Bordenave, "Stability properties of data flows in a CDMA network with macrodiversity," INRIA, Tech. Rep. RR 5257, February 2004, accepted in IEEE Inf. Theory.

[9] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," in *Proc. of IEEE INFOCOM*, March 2003, pp. 321–331.

[10] H. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Communications*, vol. 3, no. 4, 2004.

[11] F. Baccelli, B. Błaszczyszyn, and M. Karray, "Blocking rates in large CDMA networks via a spatial Erlang formula," in *Proc. of IEEE INFOCOM*, Miami, 2005.

[12] T. Bonald, S. Borst, N. Hegde, and A. Proutiere, "Wireless data performance in multi-cell scenarios," in *In Proc. of ACM Sigmetrics*, 2004.

[13] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 57–62, 1992.

[14] B. Radunovic and J.-Y. Le Boudec, "A unified framework for max-min and min-max fairness with applications," in *40th Annual Allerton Conference on Communication, Control, and Computing*, 2002.

[15] A. Maulloo, F. P. Kelly, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.

[16] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM trans. on Networking*, vol. 9, no. 5, pp. 556–567, 2000.

[17] M. Karray, "Evaluation of wireless system performance using a spatial Markov queueing process accounting for their geometry and dynamics," Ph.D. dissertation, ENST Paris, 2006.

[18] D. Tse and P. Viswanath, *Foundamentals of Wireless Communication*. Cambridge University Press, 2005.

[19] M. Karray, " "FFactor" — mean formulas for hexagonal CDMA networks with Poisson traffic," private communication, 2003.

[20] F. Baccelli and P. Brémaud, *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences*. Paris: Springer, 2003.

[21] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risk*. Wiley and Sons, 2002.