How mobility impacts the flow-level performance of wireless data systems

Thomas Bonald, Sem Borst, Alexandre Proutière

How Mobility Impacts the Flow-Level Performance of Wireless Data Systems

T. Bonald*, S.C. Borst†, A. Proutière*‡

Abstract— **The potential for exploiting rate variations to increase the capacity of wireless systems by opportunistic scheduling has been extensively studied at packet level. In the present paper, we examine how slower, mobility-induced rate variations impact performance at flow level, accounting for the random number of flows sharing the transmission resource. We identify two limit regimes, termed *fluid* and *quasi-stationary*, where the rate variations occur on an infinitely fast and an infinitely slow time scale, respectively. Using stochastic comparison techniques, we show that these limit regimes provide simple performance bounds that only depend on easily calculated load factors. Additionally, we prove that for a broad class of fading processes, performance varies monotically with the speed of the rate variations. These results are illustrated through numerical experiments, showing that the fluid and quasi-stationary bounds are remarkably tight in certain usual cases.**

I. Introduction

Next-generation wireless networks are expected to support a wide variety of high-speed data applications, in addition to conventional voice services and current low-bandwidth data services such as short messaging. The integration of these heterogeneous applications on a common transmission infrastructure raises similar challenges as in wireline integrated networks. In wireless environments, these issues are further exacerbated by interference problems, intrinsically limited bandwidth, and highly variable and unpredictable propagation characteristics. Specifically, the channel quality may vary widely among spatially distributed users due to distance-related attenuation. In addition, the channel conditions for a given user may vary dramatically over time because of fading effects.

Fading is an extremely complex physical phenomenon caused by the interaction between the propagation environment and user mobility. It emerges in diverse forms and typically spans a wide range of time scales. Fast fading arises because of multi-path propagation effects, and as the word suggests, occurs at a relatively high pace. Slow fading manifests ifself at a more macroscopic level as a result of distance-related attenuation and scattering due to obstacles and terrain conditions, and evolves over a longer time scale.

Wireless circuit-switched voice networks rely on power control mechanisms for adjusting the transmit power to combat fading and maintain a fixed transmission rate. Various data applications on the other hand, such as file transfers and Web browsing sessions, are less sensitive to packet-level delays, and do not have a stringent rate requirement. Such elastic applications are well-suited for rate control algorithms which dynamically adapt the transmission rate over time so as to match the fluctuations in channel quality. The resulting variations in the transmission rates in fact open up the possibility of scheduling data transmissions to the various users when their channel conditions are relatively favorable. While fading is considered to have a predominantly adverse impact for voice connections, it thus provides the opportunity to achieve throughput gains for elastic data transfers.

The performance gains from *opportunistic* scheduling rely on the rates varying sufficiently slowly to be tracked with reasonable accuracy, but relatively rapidly compared to the delay tolerance of the users. High-frequency fading causes estimation and prediction problems, diminishing the scope for scheduling. Slow variations cannot be harnessed, or only at the expense of compromising the delay allowance of the users. For example, typical values of the time constant in the Proportional Fair algorithm for the CDMA 1xEV-DO system [6], [11], [16] are between 10 and 1000 slots of 1.67 ms. This ensures that starvation effects cannot persist for excessive periods, but it also implies that slower variations are not exploited. In practice, relatively low-mobility scenarios tend to provide the greatest potential for scheduling gains.

While the performance of opportunistic scheduling algorithms has been thoroughly explored at packet level [2], [9], [14], [19], [23], [26], the impact of fading on flow-level performance has received remarkably little attention so far. In [8], it was shown that when fading is relatively fast compared to flow dynamics, the system may in certain cases be represented by a Processor-Sharing type model with a state-dependent service rate that accounts for the scheduling gains. This model provides explicit formulas for the distribution of the number of active flows and of the mean transfer delay. In particular, performance is *insensitive*, in the sense that these formulas only depend on the characteristics of the system through an easily computed 'load' factor. The notion of 'cell capacity', critical for dimensioning purposes, can then be defined independently of precise statistics of offered traffic [7].

In the present paper, we focus on the impact of mobility-induced fading that evolves on a slower time scale and manifests itself in the form of independent rate variations at flow level. Due to these slower rate variations, the insensitivity property is lost, and performance depends in some complicated fashion on detailed rate statistics and traffic characteristics of the system, rendering exact analysis virtually impossible. Considering these complexities, we compare the performance

*France Telecom R&D, France. ‡Corresponding author.
{thomas.bonald,alexandre.proutiere}@francetelecom.com
†CWI, The Netherlands, Sem.Borst@cwi.nl, also affiliated with Bell Laboratories, Lucent Technologies

0-7803-8355-9/04/$20.00 ©2004 IEEE.

of the system with that in two limit regimes, termed *fluid* and *quasi-stationary*, obtained when rate variations have the same instantaneous statistics, but occur on an infinitely fast and an infinitely slow time scale, respectively. Using stochastic comparison techniques, we show that the fluid and quasi-stationary regimes yield optimistic and conservative performance estimates, respectively. These estimates are very useful, since performance in the limit regimes *is* insensitive, and only depends on appropriately defined load factors, thus providing simple bounds that render the detailed statistical characteristics of the system largely irrelevant. Numerical experiments indicate that these bounds are surprisingly tight in many cases.

The above ordering results show that infinitely fast rate fluctuations yield best performance, while infinitely slow variations produce worst performance. It is tempting to conjecture that performance improves monotonically as the fading process is speeded up. We demonstrate that this is indeed the case for a broad class of Markov-type fading processes. It is worth observing that these results relate to a change in the time scale of the rate variations for *given* instantaneous rate statistics. As mentioned above, the actual transmission rates may be reduced at higher fading frequencies because of estimation and prediction problems, so for a given system, a change in the time scale will also affect the marginal rate distributions to some extent.

At a qualitative level, the finding that mobility-related rate variations improve performance resonates with the generic principle described earlier with respect to opportunistic scheduling. It also ties in with the observation in [13] that mobility increases the capacity of ad hoc wireless networks. In the present context, however, the performance improvement does not rely on opportunistic scheduling. Instead, informally stated, it arises from the fact that flow-level performance measures behave as convex functions of the rate processes.

The remainder of the paper is organized as follows. In Section II we present a detailed model description. In Section III we introduce the fluid and quasi-stationary regimes mentioned above. We establish a necessary and sufficient stability condition in Section IV. In Section V it is proved that the fluid and quasi-stationary regimes provide stochastic bounds for the performance of the actual system. For Markov-type fading processes, we demonstrate in Section VI that performance in fact monotonically varies with the time scale of rate fluctuations. In Section VII we discuss some numerical experiments performed to illustrate the analytical results. In Section VIII we make some concluding remarks.

II. Model description

Consider a single base station whose transmission power is time shared by a dynamic number of elastic flows. Each flow is represented as a 'fluid' data transfer with a variable rate that depends on the channel quality and the number of competing flows. Packet-level dynamics are implicitly represented through the way flows share the transmission resource, as explained below. Each flow is characterized by its size (in bits) and its 'feasible' transmission rate that varies relatively slowly, due notably to user mobility.

Specifically, we consider an arbitrary number K of flow classes, each class corresponding to given statistical flow size and rate variation characteristics. Class-k flows arrive as a Poisson process of rate λ_k. We denote by F_{ki} the size of the i-th arriving class-k flow and by $R_{ki}(t)$ its feasible rate at time t, corresponding to the actual transmission rate of this flow if it were the only one in the system. (For notational convenience, we define $R_{ki}(t)$ for all values of t. Note however that the i-th class-k flow may not have arrived yet or may already be completed at time t, in which case the value of $R_{ki}(t)$ is of no significance.) We assume that F_{ki} and $R_{ki}(t)$, $i = 1, 2, ..,$ are i.i.d. copies of a random variable F_k and a stationary and ergodic process $R_k(t)$, respectively. The process $R_k(t)$ is assumed to be bounded and right-continuous with left-hand limits.

Let $C_k = \mathrm{E}[R_k(0)]$ be the time-average feasible rate of a class-k flow. We define $\rho_k = \lambda_k \mathrm{E}[F_k]/C_k$ as the traffic load associated with class k and denote by $\rho = \sum_{k=1}^{K} \rho_k$ the total traffic load. It is not directly clear what the right concept of 'load' is in view of the time-varying transmission rates. In particular, the load as defined above does not coincide with the fraction of time that the base station is active. However, the results in Section IV will show that the above-defined notion does provide a correct measure of load from a stability perspective.

Assuming packet scheduling results in fair sharing at flow level, the actual transmission rate of the i-th arriving class-k flow, if present at time t, is:

$$R_{ki}(t)\frac{G(n)}{n}, \tag{1}$$

where n denotes the total number of flows present at time t. The function $G(n)$ accounts for possible throughput gains from channel-aware scheduling. In particular, the function $G(n)$ with $G(1) = 1$ is increasing in n and tends to some finite limit value G^* for $n \to \infty$, while the ratio $G(n)/n$ is decreasing in n.

Remark 1: Fair sharing trivially occurs in case of static round-robin scheduling for example, corresponding to $G(n) \equiv 1$, but it may also naturally arise in case of channel-aware scheduling. Specifically, in case $R_{ki}(t) \equiv C_k$, the model reduces to that considered in [8] for the flow-level performance of a weight-based scheduling strategy which assigns weights $w_k = 1/C_k$ to class-k users. In case the users have statistically identical normalized rate variations $Y_1, Y_2, \ldots$ at packet level, it may then be shown that $G(n) = \mathrm{E}[\max\{Y_1, \ldots, Y_n\}]$. As may further be deduced from [1], [8], [18], [24], the Proportional Fair algorithm for the CDMA 1xEV-DO system would approximately behave like the weight-based strategy, provided the exponential smoothing window is sufficiently large. In case the feasible transmission rate $R_{ki}(t)$ is (slowly) time-varying, similar arguments suggest that a weight-based strategy which assigns a dynamic weight $w_{ki}(t) = 1/R_{ki}(t)$ to the i-th class-k user, results in the actual transmission rate (1) at flow level.

Remark 2: The comparison results to be derived in Sections V and VI in fact remain valid under the even milder

assumption that the i-th class-k user receives service at rate

$$R_{ki}(t)H_k(n_1,\ldots,n_K), \tag{2}$$

where n_k denotes the number of active class-k flows and the function $H_k(\cdot)$ is decreasing in each of the n_k's. Unfortunately however, when the function $H_k(\cdot)$ is not of the form $G(n)/n$ with $n=\sum_{k=1}^K n_k$ as in (1), the fluid and quasi-stationary regimes described below prove extremely difficult to analyze.

III. Definition of fluid and quasi-stationary regimes

The flow-level model defined by (1) corresponds to a Processor-Sharing type queue where the service rate of each customer is modulated by an independent stochastic process. Considering the extreme difficulty of analyzing such a system, we introduce two limit regimes, termed *fluid* and *quasi-stationary*, where the rate processes evolve on an infinitely fast and an infinitely slow time scale, respectively. Formally, let us consider a family of systems, parameterized by $s\in(0,\infty)$, where the generic rate process for class-k flows is $R_k^{(s)}(t)\equiv R_k(st)$. Thus the parameter s represents the 'speed' of the rate process. In case $R_k(t)$ is a Markov process, the process $R_k^{(s)}(t)$ may be obtained by scaling the transition rates with s.

When the parameter s grows large, the rate process approximately averages out over the time scale of the flow dynamics. In the limit for $s\to\infty$, the variations completely vanish, and the rate process reduces to a constant, giving rise to the 'fluid' regime with $R_k^{\rm fl}(t)\equiv R_k^{(\infty)}(t)=C_k$. On the other hand, as the value of s becomes small, the fading process remains roughly constant over the time scale of the flow dynamics. In the limit for $s\to 0$, the changes completely disappear, and the rate process freezes in some initial state, yielding the 'quasi-stationary' regime with $R_k^{\rm qs}(t)\equiv R_k^{(0)}(t)=R_k(0)$, where $R_k(0)$ has the stationary marginal distribution of the process $R_k(t)$.

Accordingly, we define the class-k traffic loads in the fluid and quasi-stationary regimes: $\rho_k^{\rm fl}=\lambda_k\mathrm{E}[F_k]/C_k$ and $\rho_k^{\rm qs}=\lambda_k\mathrm{E}[F_k/R_k(0)]=\lambda_k\mathrm{E}[F_k]/C_k^{\rm qs}$, where $C_k^{\rm qs}=\mathrm{E}[1/R_k(0)]^{-1}$. Note that these load factors depend on the rate statistics only through the arithmetic and harmonic means, respectively. By Jensen's inequality, we have $\rho_k^{\rm fl}\le\rho_k^{\rm qs}$. Denote finally by $\rho^{\rm fl}\equiv\sum_k\rho_k^{\rm fl}$ and $\rho^{\rm qs}\equiv\sum_k\rho_k^{\rm qs}$ the total traffic loads in the fluid and quasi-stationary regimes, respectively. Recall that $\rho\equiv\rho^{\rm fl}$.

As mentioned earlier, the fluid and quasi-stationary regimes are particularly relevant, because their performance can be explicitly evaluated. Based on the results of [8], it can be shown that a necessary and sufficient condition for stability of the fluid (resp. quasi-stationary) regime is $\rho^{\rm fl}<G^*$ (resp. $\rho^{\rm qs}<G^*$). In case of stability, the stationary distributions $\pi^{\rm fl}$ and $\pi^{\rm qs}$ of the numbers $(n_1,\ldots,n_K)$ of on-going flows of each class in the fluid and quasi-stationary regimes depend on the traffic and rate statistics through the loads $\rho_k^{\rm fl}$ and $\rho_k^{\rm qs}$ only:

$$\pi^{\rm fl}(n_1,\ldots,n_K)=\pi^{\rm fl}(0)\frac{n!}{\prod_{i=1}^n G(i)}\prod_{k=1}^K\frac{(\rho_k^{\rm fl})^{n_k}}{n_k!},$$

$$\pi^{\rm qs}(n_1,\ldots,n_K)=\pi^{\rm qs}(0)\frac{n!}{\prod_{i=1}^n G(i)}\prod_{k=1}^K\frac{(\rho_k^{\rm qs})^{n_k}}{n_k!},$$

where $\pi^{\rm fl}(0)$ and $\pi^{\rm qs}(0)$ are determined by the normalizing condition, and $n=\sum_{k=1}^K n_k$. By Little's law, we obtain the mean response time of class-k flows:

$$\mathrm{E}[T_k]=\frac{\mathrm{E}[n_k]}{\lambda_k}.$$

Alternatively, the performance can be naturally measured in terms of *flow throughput*:

$$\gamma_k\equiv\frac{\mathrm{E}[F_k]}{\mathrm{E}[T_k]}=\frac{\rho_k C_k}{\mathrm{E}[n_k]}.$$

When $G(n)\equiv 1$, we obtain, for the fluid and quasi-stationary regimes, respectively:

$$\gamma_k^{\rm fl}=C_k(1-\rho^{\rm fl})\quad\text{and}\quad\gamma_k^{\rm qs}=C_k^{\rm qs}(1-\rho^{\rm qs}). \tag{3}$$

IV. Stability condition

We say that the system described in Section II is stable if, starting from any initial state, it converges to a finite stationary regime. It follows from the stochastic bounds to be derived in Section V that the condition $\rho^{\rm fl}<G^*$ is necessary for stability, while the condition $\rho^{\rm qs}<G^*$ is sufficient. Note that when the number of on-going flows tends to ∞, each flow stays a long time in the system so that the rate process tends to average out over the flow duration, i.e., the system behaves as in the fluid regime. Thus we expect the condition $\rho\equiv\rho^{\rm fl}<G^*$ to be both necessary and sufficient for stability. This is indeed the case in view of the following result, proved in Appendix A.

Theorem 1: If $\rho<G^*$, then the system is stable.

Remark 3: The assumption of fair sharing is crucial for the above stability condition to hold. This condition may be relaxed by giving priority to those flows with the highest feasible rates. In a high mobility scenario, the BS would then transmit to a user only when she/he is close to the BS, a strategy closely related to that considered in [13] in the context of ad hoc networks. In the present context, however, fast variations in the feasible rates are already exploited at the packet level by opportunistic scheduling and slower variations cannot be exploited without severe impact on user performance due to starvation effects (refer to Section I).

V. Comparison with fluid and quasi-stationary regimes

We now compare the performance of the system with that in the fluid and quasi-stationary regimes, using the notion of stochastic ordering (see for instance [21]).

Definition 1: (st and icx orderings) Let X and Y be two r.v.'s on $\mathbb{R}^n$. Write $X\le_{st}Y$ (resp. $X\le_{icx}Y$) if and only if $\mathrm{E}[f(X)]\le\mathrm{E}[f(Y)]$ for all increasing (resp. increasing and convex) functions $f:\mathbb{R}^n\to\mathbb{R}$ for which the previous expectations exist.

Note that these orderings are particularly relevant, since $X\le_{st}Y$ allows the comparison of the distributions of X and

Y, i.e., $\Pr[X \leq x] \geq \Pr[Y \leq x]$ for all x, whereas $X \leq_{icx} Y$ implies all moments of X are less than the respective moments of Y.

Assume that the system is empty at time 0 and denote by $N_k(t)$ the number of active class-k flows at time t. For $i = 1, \ldots, N_k(t)$, let $F_{ki}(t)$ be the remaining size of active class-k flow i at time t. We define the total workload at time t as:

$$W(t) = \sum_{k=1}^{K} \sum_{i=1}^{N_k(t)} \frac{F_{ki}(t)}{C_k}.$$

Theorem 2 below, proved in Appendix B, states that performance improves (resp. deteriorates) in terms of the number of active flows, the workload and the response time T of an arbitrary flow, when the rate processes of some flows satisfying Assumption 1 below are replaced by the corresponding fluid (resp. quasi-stationary) versions as described in Section III.

Assumption 1: The cumulative distribution function (c.d.f.) $P(\cdot) = \Pr[F \leq \cdot]$ associated with the random flow size F is concave.

Note that Assumption 1 is satisfied by a broad class of distributions, including exponential, hyper-exponential or Weibull. In particular, it is possible to represent the highly variable flow size distribution of typical data networks.

Theorem 2: We have, for all $k = 1, \ldots, K$,

$$W^{\text{fl}}(t) \leq_{icx} W(t) \leq_{icx} W^{\text{qs}}(t), \tag{4}$$

$$N_k^{\text{fl}}(t) \leq_{st} N_k(t) \leq_{st} N_k^{\text{qs}}(t), \tag{5}$$

$$T^{\text{fl}} \leq_{st} T \leq_{st} T^{\text{qs}}, \tag{6}$$

where the superscript $^{\text{fl}}$ (resp. $^{\text{qs}}$) refers to the system where the rate processes of some set of flows satisfying Assumption 1, are replaced by the corresponding fluid (resp. quasi-stationary) processes.

The above comparison results are also valid when the system is in equilibrium. Denote by $W(\infty)$, $N_k(\infty)$ and $T_k(\infty)$ the workload, the number of active class-k flows and the response time of class-k flows in steady state, respectively. We deduce the inequalities (8) and (9) in the next corollary from Theorem 2 and the stability of the st-order by limits [21]. The inequality (7) results from (4) and a classical monotonicity property of the Loynes' construction as explained in [3, page 281].

Corollary 1: Let $\mathcal{L} \subset \{1, \ldots, K\}$ be an arbitrary subset of classes that satisfy Assumption 1. We have, for all $k = 1, \ldots, K$:

$$W^{\text{fl}}(\infty) \leq_{icx} W(\infty) \leq_{icx} W^{\text{qs}}(\infty), \tag{7}$$

$$N_k^{\text{fl}}(\infty) \leq_{st} N_k(\infty) \leq_{st} N_k^{\text{qs}}(\infty), \tag{8}$$

$$T_k^{\text{fl}}(\infty) \leq_{st} T_k(\infty) \leq_{st} T_k^{\text{qs}}(\infty), \tag{9}$$

where the superscript $^{\text{fl}}$ (resp. $^{\text{qs}}$) refers to the system where the rate processes of the flows of the classes in $\mathcal{L}$ are replaced by the corresponding fluid (resp. quasi-stationary) processes.

Note that the above fluid and quasi-stationary regimes correspond to those defined in Section III when $\mathcal{L} = \{1, \ldots, K\}$ in which case they provide tractable upper and lower performance bounds.

VI. Impact of the speed of rate variations

In this section, we investigate how performance varies with the time scale of the rate processes. In order to do so, we suppose that the processes $R_k(t)$ for some users are replaced by processes $R_k^{(s)}(t) \equiv R_k(st)$ for some constant $s > 1$. The constant s may be interpreted as an acceleration factor. Although one might conjecture that performance improves when the rate process is speeded up, this result does not hold in certain very specific cases [22]. However, the monotonity property can be established when the rate process satisfies the following assumption.

Assumption 2: The rate process is a homogeneous stationary Markov process. The transition kernels Q and Q_r of the Markov process and of the corresponding time-reversed Markov process are $\leq_{st}$-monotone. Recall that Q is $\leq_{st}$-monotone if and only if, for all increasing functions f, the function $x \mapsto \int f(t)Q(x, dt)$ is also increasing [21].

Assumption 2 is satisfied by a broad class of processes, including birth-death processes and Markov processes with a discrete state space and a generator $Q = (q_{ij})$ such that q_{ij} does not depend on i [5].

The next theorem, proved in Appendix B, states that performance improves when the rate processes of some set of flows satisfying Assumptions 1 and 2 are accelerated. Speeding up some users (or equivalently their rate process) improves the performance for all flows.

Theorem 3: We have, for all $s > 1$ and all $k = 1, \ldots, K$:

$$W^{(s)}(t) \leq_{icx} W(t), \tag{10}$$

$$N_k^{(s)}(t) \leq_{st} N_k(t), \tag{11}$$

$$T^{(s)} \leq_{st} T, \tag{12}$$

where the superscript $^{(s)}$ refers to the system where the rate processes of some set of flows satisfying Assumptions 1 and 2, are speeded up by a factor s.

The next corollary presents the counterpart of Corollary 1.

Corollary 2: Let $\mathcal{L} \subset \{1, \ldots, K\}$ be an arbitrary subset of classes that satisfy Assumptions 1 and 2. We have, for all $s > 1$ and all $k = 1, \ldots, K$:

$$W^{(s)}(\infty) \leq_{icx} W(\infty), \tag{13}$$

$$N_k^{(s)}(\infty) \leq_{st} N_k(\infty), \tag{14}$$

$$T_k^{(s)}(\infty) \leq_{st} T_k(\infty), \tag{15}$$

where the superscript $^{(s)}$ refers to the system where the rate processes of the flows of the classes in $\mathcal{L}$ are sped up by a factor s.

VII. Numerical Experiments

We now present some numerical experiments to illustrate the theoretical results presented above. The feasible rate of a user is a complex function depending on both fast and slow fading. For the sake of simplicity in simulations, we ignore fast fading. Slow fading may be viewed as the result of two different phenomena: shadowing and variations in path loss. Thus, we assume that the feasible rate $R(t)$ of a user behaves as:

$$R(t) \propto G_s(t) \times \Gamma(t),$$

where $G_s(t)$ and $\Gamma(t)$ denote the shadowing component (of unit mean) and the path loss, respectively. Shadowing typically arises when the variation in the distance to the base station is of the order of 10-100 m, depending on the propagation environment. Empirical studies suggest that shadowing has a log-normal distribution, with standard deviation between 5 and 12 dB. Path loss usually varies over larger distances. Here we assume that $\Gamma(t)$ is proportional to $r(t)^{-\alpha}$, where $r(t)$ denotes the distance to the base station and α is the path loss exponent.

Based on these observations, we consider the following two mobility scenarios:

- *Low* mobility, where the typical distance covered by a moving user during the flow transfer is less than 100 m. Variations in the feasible rate are then mainly due to shadowing and result from variations of 5 to 12 dB in the received signal.
- *High* mobility, where a user can move across the entire cell during the flow transfer. In this case, fluctuations in the feasible rate are mainly due to variations in path loss.

In the CDMA 1xEV-DO system, 11 feasible rates are defined, $c_0 > c_1 > \ldots > c_{10}$, with corresponding target signal to interference-plus-noise ratios (SINR). In case of a constant shadowing component $G_s(t) \equiv 1$, these rates correspond to a set of concentric rings of radius $r_0 < r_1 < \ldots < r_{10}$, such that when $r(t) \in (r_{j-1}, r_j)$, the feasible rate is c_j (with the convention $r_{-1} \equiv 0$) [7]. We give the ring radius (normalized so that $r_0 \equiv 1$) corresponding to a path loss exponent $\alpha = 4$ in Table I.

Ring j	Rate c_j (Kbit/s)	Radius r_j ($\alpha = 4$)	SINR (dB)
0	2457.6	1	9.5
1	1843.2	1.07	7.2
2	1228.8	1.19	3.0
3	921.6	1.28	1.3
4	614.4	1.41	−1.0
5	307.2	1.68	−4.0
6	204.8	1.86	−5.7
7	153.6	2.00	−6.5
8	102.6	2.21	−8.5
9	76.8	2.37	−9.5
10	38.4	2.82	−12.5

TABLE I
RATES, RING RADIUS AND SINR FOR CONSTANT SHADOWING

In the following, we consider a circular cell of external radius $R = r_L$ corresponding to $L + 1$ rings. Flows arrive uniformly in the cell according to a Poisson process of intensity λ. The probability p_j that a new flow starts its service in ring j is proportional to the surface of this ring, i.e., $p_j = (r_j^2 - r_{j-1}^2)/R^2$. The flow throughputs in the limit regimes are given by (3). Simulation results are obtained for exponentially distributed flow sizes of unit mean and Markov rate processes with values in $\{c_0, c_1, \ldots, c_L\}$. We make the natural assumption that the rate process can only jump between adjacent states, so that for each class, the Markov rate process is a birth-death process. Note that Assumptions 1 and 2 are satisfied.

A. Low mobility

In the low-mobility scenario, the feasible rate of a user typically evolves in a set of 2 to 5 consecutive rates, roughly corresponding to SINR variations of 5 to 12 dB (see Table I). Rather than fitting a log-normal distribution, we simply assume that the feasible rate of each user takes a fixed number of values, a, and that all transitions rates of the corresponding birth-death process are equal. We performed simulations not reported here to verify that performance depends on shadowing mainly through its amplitude and not on its precise distribution. We consider a cell of radius $R = 1.86$ (thus $L = 6$) and evaluate performance in the following two cases:

- Shadowing with low amplitude ($a = 3$). There are 5 user classes. Class-k users are located in ring k, $k = 1, \ldots, 5$, and their feasible rates are c_{k-1}, c_k, c_{k+1} with corresponding marginal probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$.
- Shadowing with high amplitude ($a = 5$). There are 3 user classes. Class-k users are located in ring k, $k = 2, \ldots, 4$, and their feasible rates are $c_{k-2}, c_{k-1}, c_k, c_{k+1}, c_{k+2}$ with corresponding marginal probabilities $\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}$.

Figure 1 presents the throughput of flows of classes 1 and 5 as a function of cell load in case of shadowing with low amplitude ($a = 3$) and with different values of the speed s. In Figure 2, we present the throughput of flows of classes 1 and 3 in case of shadowing with high amplitude ($a = 5$).

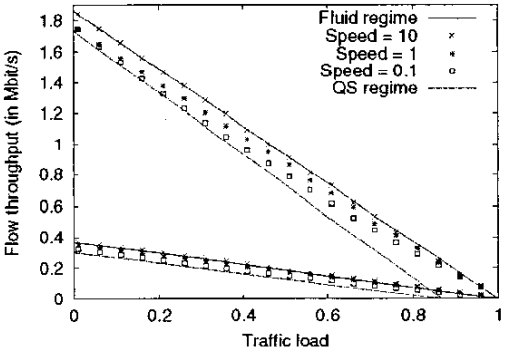

Fig. 1. Throughput of flows of class 1 (upper curves) and class 5 (lower curves) in case of shadowing of low amplitude ($a = 3$).

As expected in view of Corollaries 1 and 2, the fluid and quasi-stationary regimes provide optimistic and conservative estimates of the throughput, respectively, and speeding up the rate processes improves performance. Further observe that the

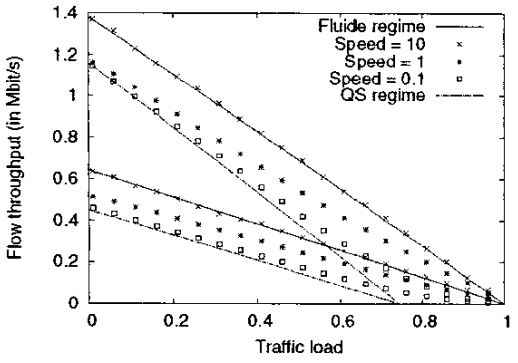

Fig. 2. Throughput of flows of class 1 (upper curves) and class 3 (lower curves) in case of shadowing of high amplitude ($a = 5$).

limit regimes only differ significantly in case of shadowing with high amplitude.

B. High mobility

We now consider a high-mobility scenario where the variations in path loss cannot be neglected. We assume that a fraction β of the users move across the entire cell while the others are static. We do not account for shadowing, i.e., $G_s \equiv 1$. There are $K = L + 2$ user classes:

- Class-k users, for $k = 0, \ldots, L$, are static in ring k, i.e., $R_k(t) \equiv C_k = c_k$ for all t. The load associated with this class is $\rho_k = (1 - \beta)\lambda p_k / C_k$.
- Class-$(L+1)$ users move in rings $0, \ldots, L$ according to a birth-death process with marginal distribution $p_0, \ldots, p_L$, corresponding to isotropic motion in the cell, so that $C_{L+1} = \sum_k p_k c_k$. The load associated with this class is $\rho_{L+1} = \beta\lambda / C_{L+1}$.

Figure 3 gives, for a cell of radius $R = 2$ (thus $L = 7$) where all users move ($\beta = 1$), the flow throughput as a function of total traffic load for different values of the speed s. The impact of speed on flow throughput for fixed load $\rho = 0.5$ is shown in Figure 4. Figure 5 is the analog of Figure 3 for a cell of radius $R = 1.19$ (thus $L = 2$). Note that for large variations in the feasible rate (see Figure 3), performance is very sensitive to speed, whereas for limited variations (see Figure 5), the fluid and quasi-stationary bounds are very close, indicating that performance is approximately insensitive.

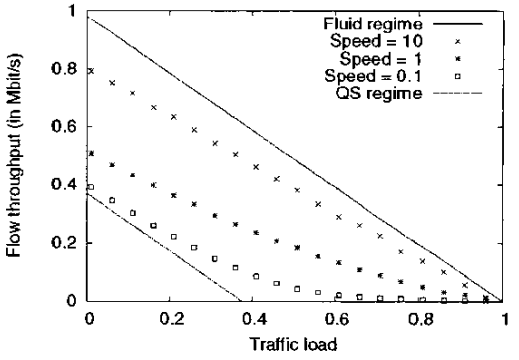

Fig. 3. Flow throughput as a function of traffic load when all users move in a cell of radius $R = 2$.

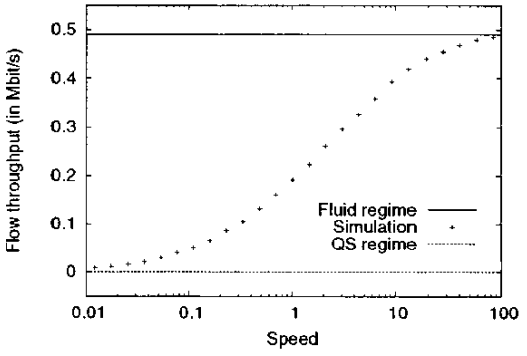

Fig. 4. Flow throughput as a function of speed for a fixed traffic load $\rho = 0.5$ when all users move in a cell of radius $R = 2$.

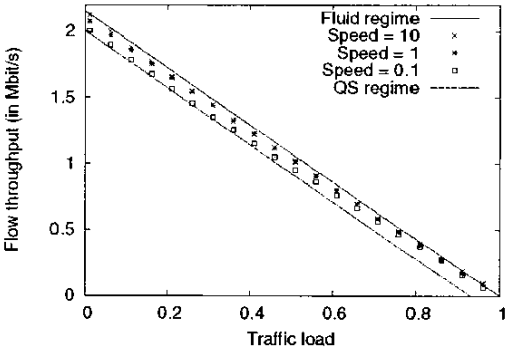

Fig. 5. Flows throughput when all users move in a cell of radius $R = 1.19$.

Figure 6 gives the flow throughput of static users in ring 0 and of moving users, when the proportions of static and moving users are the same ($\beta = 0.5$).

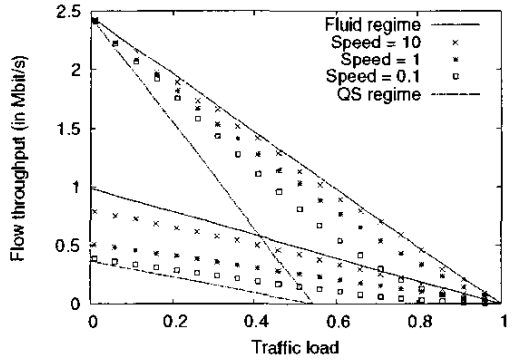

Fig. 6. Flow throughput of static users in ring 0 (upper curves) and of moving users (lower curves) in a cell of radius $R = 2$.

The numerical results suggest that performance is sensitive to the speed of the fading process only when variations in the feasible rates of those users representing a significant part of the total traffic load are of high amplitude. When the amplitude is low, performance is almost insensitive, i.e., essentially depending on the traffic and fading statistics through the traffic load of each class only. In this case, the quasi-stationary regime provides an accurate conservative estimate of flow-level performance.

VIII. Conclusion

We have examined the impact of slow, mobility-induced rate variations on the flow-level performance of a wireless data system. We have compared the performance of the system with that in two limit regimes, termed fluid and quasi-stationary, where rate variations occur on an infinitely fast and an infinitely slow time scale, respectively. The fluid and quasi-stationary regimes provide explicit performance estimates, which are provably optimistic and conservative, respectively. Besides, the performance of the limit regimes is insensitive, and only depends on an appropriately defined load factor, thus yielding bounds that only involve simple first-order system parameters For a broad class of Markov-type fading processes, we further proved that performance varies monotonically with the time scale of the rate variations.

At a qualitative level, the finding that mobility-related rate variations improve performance resembles the generic principle underlying opportunistic scheduling. In the present context, however, the performance improvement does not rely on opportunistic scheduling. Instead, informally stated, it arises from the fact that flow-level performance measures behave as convex functions of the rate processes.

From a practical perspective, when the traffic load generated by those users with large rate variations is limited, the quasi-stationary regime provides an accurate conservative estimate of flow-level performance. This allows the development of simple dimensioning rules *as if* users were static, as derived in [7]. In cases where moving users with large rate variations represent a significant part of the total traffic load, performance becomes sensitive to the precise traffic and fading statistics. It may then be necessary to take mobility and shadowing effects into account.

Note finally that the positive impact of mobility relies here on the assumption of perfect rate predictions. This is reflected in the model by the fact that the marginal distributions of the feasible rates do not depend on the time scale of rate variations. It would be very interesting to study the extent to which the estimation and prediction problems due to high-frequency fading counterbalance the performance improvements established in the present paper.

Appendix

A. Stability condition

Proof of Theorem 1: We prove the result for Cox flow size distributions which are known to form a dense subset of the set of all distributions with non-negative support. Specifically, we assume that class-k flows have i.i.d. exponential sizes of mean $1/\mu_k$, and generate a new class-l flow with probability p_{kl} when completed. By creating additional classes and dividing each random flow size into a random number of exponential phases, the model is then sufficiently general to cover any flow size distribution. The total traffic load is given by:

$$\rho = \lambda(I - P)^{-1}(\mu C)^{-1},$$

where $\lambda = (\lambda_k)$ is a row vector, I is the identity matrix, $P = (p_{kl})$, and $\mu = (\mu_k)$, $C = (C_k)$ are diagonal matrices. Similarly, we assume that the rate process $R_{ki}(t)$ is a function of a finite-state Markov process $\sigma_{ki}(t)$. By increasing the number of states, such a Markov process can approximate any stationary and ergodic process.

The stochastic process $\{N(t), \sigma(t)\}$ where $N(t)$ and $\sigma(t)$ denote the row vectors $(N_k(t))$ and $(\sigma_{ki}(t))$, $k = 1, \ldots, K$, $i = 1, \ldots, N_k(t)$, respectively, is an irreducible Markov process. Define the workload at time t as:

$$W(t) = |N(t)(I - P)^{-1}(\mu C)^{-1}|.$$

Assume that $\rho < G^*$ and let $t_0 \equiv 1/(G^* - \rho)$. For any sequence of initial states $\{N^{(j)}(0), \sigma^{(j)}(0)\}_j$ with

$$\lim_{j\to\infty} \frac{W^{(j)}(0)}{j} = 1, \tag{16}$$

we will prove that the sequence of workload processes $\{W^{(j)}(t)\}_j$ satisfies for any $t < t_0$:

$$\lim_{j\to\infty} \frac{\mathrm{E}\left[W^{(j)}(jt)\right]}{j} = 1 - (G^* - \rho)t.$$

As the workload defines a Lyapunov function for the Markov process $\{N(t), \sigma(t)\}$, the proof then follows from Foster's stability criterion [20].

Denote by $A(t)$, $B(t)$ and $D(t)$, respectively, the row vectors of the number of exogenous arrivals, endogenous arrivals and departures of class-k flows up to time t, $k = 1, \ldots, K$. We have:

$$N(t) = N(0) + A(t) + B(t) - D(t). \tag{17}$$

Let $D'(t)$ be the row vector of the maximum number of departures of class-k flows up to time t, thus assuming these flows are served at the rate $S_k \equiv G^* \sup_t R_k(t)$. Similarly, let $B'(t)$ be the row vector of the maximum number of endogenous arrivals of class-k flows up to time t. We get:

$$W(t) \leq W(0) + |(A(t) + B'(t))(I-P)^{-1}(\mu C)^{-1}|. \quad (18)$$

Denoting by e the row vector $(1,\ldots,1)$ and by S the diagonal matrix (S_k), it follows from the strong law of large numbers that:

$$\frac{A(jt)}{j} \overset{\text{a.s.}}{\to} \lambda t, \quad \frac{D'(jt)}{j} \overset{\text{a.s.}}{\to} e\mu St, \quad \frac{B'(jt)}{j} \overset{\text{a.s.}}{\to} e\mu PSt,$$

when $j \to \infty$. In particular, there exists for any sequence of initial states $\{N^{(j)}(0), \sigma^{(j)}(0)\}_j$ satisfying (16) a subsequence denoted by indices j' such that for any $t < t_0$:

$$\frac{W^{(j')}(j't)}{j'} \overset{\text{a.s.}}{\to} \bar{W}(t).$$

The function $\bar{W}(\cdot)$ is usually referred to as a 'fluid limit' due to the time-space scaling [12]. As $D(t) \leq |D'(t)|$ and $B(t) \leq B'(t)$, there exists a subsequence of j', denoted by indices j'', and continuous functions $\bar{B}$ and $\bar{D}$ such that for any $t < t_0$:

$$\frac{B^{(j'')}(j''t)}{j''} \overset{\text{a.s.}}{\to} \bar{B}(t), \quad \frac{D^{(j'')}(j''t)}{j''} \overset{\text{a.s.}}{\to} \bar{D}(t) \text{ when } j'' \to \infty.$$

It then follows from (17) that:

$$\frac{N^{(j'')}(j''t)}{j''} \overset{\text{a.s.}}{\to} \bar{N}(t), \text{ when } j'' \to \infty,$$

where $\bar{N}$ is the continuous function given by:

$$\bar{N}(t) = \bar{N}(0) + \lambda t + \bar{B}(t) - \bar{D}(t). \quad (19)$$

Now as flow sizes are i.i.d. exponential, the number of departures of class-k flows during the time interval $[u,v]$ satisfies:

$$D_k(v) - D_k(u) \leq \sum_{i=1}^{M_k(u,v)} 1\left\{F_{ki} \leq \frac{G^*}{|m(u,v)|}\int_u^v R_{ki}(t)dt\right\},$$

where $M(u,v) = N(u) + A(v) - A(u) + B(v) - B(u)$ and $m(u,v) = N(u) - D(v) + D(u)$ correspond to the row vectors of the maximum and the minimum number of active class-k flows during $[u,v]$, respectively. Analogously, we define $\bar{M}(u,v) = \bar{N}(u) + \lambda(v-u) + \bar{B}(v) - \bar{B}(u)$ and $\bar{m}(u,v) = \bar{N}(u) - \bar{D}(v) + \bar{D}(u)$. For any u such that $|\bar{N}(u)| > 0$, we have $|m(u,v)| > 0$ for sufficiently small $v > u$. For any $\varepsilon > 0$, there exists l such that for all $j'' \geq l$,

$$\frac{1}{j''}|N^{(j'')}(j''u) - D^{(j'')}(j''v) + D^{(j'')}(j''u)| \geq |m(u,v)|(1-\varepsilon) \quad \text{a.s.}$$

and by the ergodicity of the process $R_k(t)$,

$$\Pr\left[\sup_{j\geq l}\left\{\frac{1}{j(v-u)}\int_{ju}^{jv} R_k(t)dt\right\} > C_k(1+\varepsilon)\right] \leq \varepsilon.$$

Writing

$$D_k(v) - D_k(u) \leq \sum_{i=1}^{M_k(u,v)} 1\left\{F_{ki} \leq \frac{G^*C_k(1+\varepsilon)(v-u)}{|m(u,v)|}\right\} + \sum_{i=1}^{M_k(u,v)} 1\left\{\frac{1}{v-u}\int_u^v R_{ki}(t)dt > C_k(1+\varepsilon)\right\},$$

it follows from the strong law of large numbers that:

$$\bar{D}_k(v) - \bar{D}_k(u) \leq \bar{M}_k(u,v) \times \left(\Pr\left[F_k \leq \frac{G^*C_k(1+\varepsilon)(v-u)}{|\bar{m}(u,v)|(1-\varepsilon)}\right] + \varepsilon\right).$$

Since this inequality holds for any $\varepsilon > 0$, we deduce:

$$\bar{D}_k(v) - \bar{D}_k(u) \leq \bar{M}_k(u,v)\Pr\left[F_k \leq \frac{G^*C_k(v-u)}{|\bar{m}(u,v)|}\right].$$

Similarly, using the fact that

$$D_k(v) - D_k(u) \geq \sum_{i=1}^{N_k(u)} 1\left\{F_{ki} \leq \frac{G(|m(u,v)|)}{|M(u,v)|}\int_u^v R_{ki}(t)dt\right\},$$

we obtain:

$$\bar{D}_k(v) - \bar{D}_k(u) \geq \bar{N}_k(u)\Pr\left[F_k \leq \frac{G^*C_k(v-u)}{|\bar{M}(u,v)|}\right].$$

Using the latter inequalities and the fact that $\bar{M}(u,v)$ and $\bar{m}(u,v)$ tend to $\bar{N}(u)$ when v tends to u, we deduce:

$$\frac{d\bar{D}}{dt}(u) = G^*\frac{\bar{N}(u)}{|\bar{N}(u)|}\mu C.$$

Analogously, one may prove that:

$$\frac{d\bar{B}}{dt}(u) = G^*\frac{\bar{N}(u)}{|\bar{N}(u)|}\mu CP.$$

Now it follows from (19) that for any $t < t_0$ such that $|\bar{N}(t)| > 0$:

$$\frac{d\bar{N}}{dt}(t) = \lambda - G^*\frac{\bar{N}(t)}{|\bar{N}(t)|}\mu C(I-P).$$

Using the fact that:

$$\bar{W}(t) = |\bar{N}(t)(I-P)^{-1}(\mu C)^{-1}|,$$

we obtain for any $t < t_0$ such that $\bar{W}(t) > 0$,

$$\frac{d\bar{W}}{dt}(t) = \rho - G^*.$$

We know that $\bar{W}(0) = 1$ in view of (16), so that $\bar{W}(t) = 1 - (G^* - \rho)t$ for any $t < t_0$. In particular, the function $\bar{W}(\cdot)$ is uniquely defined for any $t < t_0$ and

$$\frac{W^{(j)}(jt)}{j} \overset{\text{a.s.}}{\to} \bar{W}(t) \quad \text{when } j \to \infty.$$

Finally, the sequence of r.v.'s $\{\frac{\mathrm{E}[W^{(j)}(jt)]}{j}\}_j$ is uniformly integrable in view of (18), so that:

$$\lim_{j\to\infty}\frac{\mathrm{E}\left[W^{(j)}(jt)\right]}{j} = \bar{W}(t).$$

B. Stochastic comparisons

1) Proof of Theorem 2: We first prove (4) and (5) for the following slotted system. The interval $(0,t)$ is divided into L slots such that the feasible rate of each flow is constant during each slot and equal to the feasible rate at the beginning of the slot. We also assume that when a flow is present at the beginning of a slot, it remains in the system during the entire slot. The proof of (4) and (5) for a non-slotted system then follows from the fact that for $L = 2^p$, $p \geq 1$, the workload and the number of class-k flows in a slotted system where the feasible rate of a flow during a slot is fixed at its maximum in the slot (respectively, its minimum) converge monotonically to $W(t)$ and $N_k(t)$, respectively, when p tends to ∞.

The proof is based on the notion of supermodular functions (see, e.g., [21]) and on Lorentz inequality [25]:

Definition 2: (Supermodular functions) $f : \mathbb{R}^n \to \mathbb{R}$ is supermodular if and only if for all $x, y \in \mathbb{R}^n$, $f(x \vee y) + f(x \wedge y) \geq f(x) + f(y)$, where $(x \vee y)_i = x_i \vee y_i$ and $(x \wedge y)_i = x_i \wedge y_i$, for all $i = 1, \ldots, n$.

Lemma 1: (Lorentz inequality) Let $Z_1, \ldots, Z_n$ be identically distributed r.v.'s. For all supermodular functions f, $\mathrm{E}[f(Z_1, \ldots, Z_n)] \leq \mathrm{E}[f(Z_1, \ldots, Z_1)]$.

Consider an arbitrary flow, say flow 1, that arrived at the beginning of slot $l \in \{1, \ldots, L\}$. Assume without loss of generality that flow 1 is of class 1. We fix the arrival process up to slot L, the rate processes of all flows except flow 1, and the sizes of all flows except flow 1. We denote by x_j the feasible rate of flow 1 during slot j, by F_1 the size of flow 1, and by E_1 the expectation with respect to the random variable F_1. Let W_L and $N_{k,L}$ be the workload and the number of class-k flows at the end of slot L, respectively. The key result is:

Lemma 2: For all $k = 1, \ldots, K$, $\mathrm{E}_1[f(W_L)]$ and $\mathrm{E}_1[g(N_{k,L})]$ are supermodular and convex functions of $(x_l, \ldots, x_L)$, for all increasing and convex functions f and all increasing functions g.

Proof. Let n_j be the number of flows present during slot j, assuming that flow 1 is present during this slot. Note that these numbers do not depend on $(x_l, \ldots, x_L)$ nor on F_1. Let W_L^1 be the workload due to flow 1 at the end of slot L, i.e., the remaining size of flow 1 divided by C_1. As the transmission rate of flow 1 in slot j is $x_j G(n_j)/n_j$, we have:

$$W_L^1 = \frac{1}{C_1} \max\left(0, F_1 - \sum_{j=l}^{L} x_j \frac{G(n_j)}{n_j}\right),$$

which, composed with an increasing and convex function, is known to be supermodular and convex in $(x_l, \ldots, x_L)$ [4]. Now let w_j and $n_{k,j}$ be the workload and the number of active class-k flows at the end of slot L, respectively, assuming flow 1 leaves the system at the end of slot j. Note that these quantities do not depend on $(x_l, \ldots, x_L)$ nor on F_1. We have:

- If $F_1 \leq x_l G(n_l)/n_l$, $W_L = w_l$ and $N_{k,L} = n_{k,l}$;
- for $l' = l+1, \ldots, L-1$, if $\sum_{j=l}^{l'-1} x_j G(n_j)/n_j < F_1 \leq \sum_{j=l}^{l'} x_j G(n_j)/n_j$, $W_L = w_{l'}$ and $N_{k,L} = n_{k,l'}$;
- if $F_1 > \sum_{j=l}^{L-1} x_j G(n_j)/n_j$, $W_L = w_L + W_L^1$ and $N_{k,L} = n_{k,L}$.

Averaging with respect to the size of flow 1, we obtain for all increasing and convex functions f:

$$\mathrm{E}_1[f(W_L)] = (f(w_l) - f(w_{l+1}))P_1\left(x_l \frac{G(n_l)}{n_l}\right) + \ldots + (f(w_{L-1}) - f(w_L))P_1\left(\sum_{j=l}^{L-1} x_j \frac{G(n_j)}{n_j}\right) + f(w_L)P_1\left(\sum_{j=l}^{L-1} x_j \frac{G(n_j)}{n_j}\right) + \mathrm{E}_1\left[f(w_L + W_L^1)1_{\{F_1 > \sum_{j=l}^{L-1} x_j \frac{G(n_j)}{n_j}\}}\right],$$

where P_1 denotes the c.d.f. of F_1. Note that the sum of the last two terms in the latter expression is simply equal to $\mathrm{E}_1[f(w_L + W_l^1)]$, which is a supermodular and convex function of $(x_l, \ldots, x_L)$. In addition, it follows from Assumption 1 that for all $m = l, \ldots, L-1$, the function

$$(x_l, \ldots, x_L) \mapsto -P_1\left(\sum_{j=l}^{m} x_j \frac{G(n_j)}{n_j}\right),$$

as the composition of an affine function and a convex function, is supermodular and convex [4]. As $G(n)/n$ decreases in n, we have $w_l \leq \ldots \leq w_L$ so that $\mathrm{E}_1[f(W_L)]$, as the sum of supermodular and convex functions, is supermodular and convex.

Similarly, we have for all increasing functions g:

$$\mathrm{E}_1[g(N_{k,L})] = (g(n_{k,l}) - g(n_{k,l+1}))P_1\left(x_l \frac{G(n_l)}{n_l}\right) + \ldots + (g(n_{k,L-1}) - g(n_{k,L}))P_1\left(\sum_{j=1}^{L-1} x_j \frac{G(n_j)}{n_j}\right) + g(n_{k,L}).$$

As $G(n)/n$ decreases in n, we have $n_{k,l} \leq \ldots \leq n_{k,L}$. Thus $\mathrm{E}_1[g(N_{k,L})]$, as the sum of supermodular and convex functions, is supermodular and convex. □

Now, for any function f, we have:

$$\mathrm{E}\left[f(W_L^{\mathrm{fl}1})\right] = \mathrm{E}\left[\mathrm{E}_1[f(W_L)](\mathrm{E}[x_l], \ldots, \mathrm{E}[x_l])\right],$$
$$\mathrm{E}[f(W_L)] = \mathrm{E}\left[\mathrm{E}_1[f(W_L)](x_l, \ldots, x_L)\right],$$
$$\mathrm{E}\left[f(W_L^{\mathrm{qs}1})\right] = \mathrm{E}\left[\mathrm{E}_1[f(W_L)](x_l, \ldots, x_l)\right],$$

where $^{\mathrm{fl}}$ (resp. $^{\mathrm{qs}1}$) denotes the fluid (resp. quasi-stationary) regime with respect to flow 1. Similar relations hold for the number of class-k flows. Using the independence of the rate processes, we deduce from Lemma 2 and the fact that $(\mathrm{E}[x_l], \ldots, \mathrm{E}[x_l]) \leq_{icx} (x_l, \ldots, x_L)$ [3] that for all increasing and convex functions f, all increasing functions g, and all $k = 1, \ldots, K$:

$$\mathrm{E}\left[f(W_L^{\mathrm{fl}1})\right] \leq \mathrm{E}[f(W_L)], \ \mathrm{E}\left[g(N_{k,L}^{\mathrm{fl}1})\right] \leq \mathrm{E}[g(N_{k,L})],$$

Similarly, using the independence of the rate processes, it follows from Lemma 2 and Lorentz inequality that:

$$\mathrm{E}[f(W_L)] \leq \mathrm{E}\left[f(W_L^{\mathrm{qs1}})\right],\ \mathrm{E}[g(N_{k,L})] \leq \mathrm{E}\left[g(N_{k,L}^{\mathrm{qs1}})\right].$$

We obtain (4) and (5) by applying successively the same reasoning to an arbitrary set of flows satisfying Assumption 1.

We now prove inequality (6). Let T_L be the time spent by an arbitrary flow in the slotted system up to slot L. We prove exactly as in Lemma 2 that that for all increasing functions g, $\mathrm{E}_1[g(T_L)]$ is a supermodular and convex function of $(x_1, \ldots, x_L)$. We deduce as above that:

$$\mathrm{E}\left[g(T_L^{\mathrm{fl}})\right] \leq \mathrm{E}[g(T_L)] \leq \mathrm{E}[g(T_L^{\mathrm{qs}})],$$

and by letting L tend to ∞,

$$\mathrm{E}\left[g(T(t)^{\mathrm{fl}})\right] \leq \mathrm{E}[g(T(t))] \leq \mathrm{E}[g(T(t)^{\mathrm{qs}})],$$

where $T(t)$ denotes the time spent by an arbitrary flow in the non-slotted system up to time t. We obtain (6) by letting t tend to ∞.

2) Proof of Theorem 3: The proof is based on the following recent result by Hu & Pan [15] (a similar result had been used by Chang, Chao & Pinedo [10]).

Lemma 3: Let $\{Z(t), t \in \mathbb{R}\}$ be a process satisfying Assumption 2. For all integers n, all $(r_1, \ldots, r_n)$ and $(s_1, \ldots, s_n)$ such that $r_1 \leq s_1$ and $r_i - r_{i-1} \leq s_i - s_{i-1}$ for $i = 2, \ldots, n$, we have:

$$\mathrm{E}[f(Z(r_1), \ldots, Z(r_n))] \geq \mathrm{E}[f(Z(s_1), \ldots, Z(s_n))]$$

for all supermodular functions $f : \mathbb{R}^n \to \mathbb{R}$ such that the previous expectations exist.

We give the proof of (10) and (11) only (the proof of (12) is then similar to that of (6)). Again, it is sufficient to prove these inequalities for a slotted system. Consider some arbitrary flow 1 satisfying Assumptions 1 and 2. Denote by $x_j = R_1(t \times j/L)$ (resp. $x_j^{(s)} = R_1(st \times j/L)$) the rate of flow 1 during slot j in the actual system (resp. when its rate process is accelerated by a factor $s > 1$). It follows from the independence of the rate processes, Lemmas 2 and 3, that for all increasing and convex functions f, all increasing functions g, and all $k = 1, \ldots, K$:

$$\mathrm{E}\left[f(W_L^{(s)1})\right] \leq \mathrm{E}[f(W_L)],\ \mathrm{E}\left[g(N_{k,L}^{(s)1})\right] \leq \mathrm{E}[g(N_{k,L})],$$

where $^{(s)1}$ refers to the system where the rate process of flow 1 is accelerated by a factor s. We obtain (10) and (11) by applying successively the same reasoning to an arbitrary set of flows satisfying Assumptions 1 and 2.

REFERENCES

[1] Agrawal, R., Subramanian, V. (2002). Optimality of certain channel-aware scheduling policies. In: *Proc. 40th Annual Allerton Conf. Commun. Control., Comp.*, 1532–1541.

[2] Andrews, D.M., Kumaran, K., Ramanan, K., Stolyar, A.L., Vijayakumar, R., Whiting, P.A.. Scheduling in a queueing system with asynchronously varying service rates. To appear in *Prob. Eng. Inf. Sc.*.

[3] Baccelli, F., Brémaud, P. (2003). *Elements of Queueing Theory*, Springer Verlag.

[4] Bäuerle, N. (1997). Inequalities for stochastic models via supermodular orderings. *Commun. Stat. – Stoc. Models* **13**, 181–201.

[5] Bäuerle, N., Rolski, T. (1998). A monotonicity result for the workload in Markovian modulated queues. *J. Appl. Prob.* **35**, 741–747.

[6] Bender, P., Black, P., Grob, M., Padovani, R., Sindhushayana, N., Viterbi, A. (2000). CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Commun. Mag.* **38** (7), 70–77.

[7] Bonald, T., Proutière, A. (2003). Wireless downlink channels: user performance and cell dimensioning. In: *Proc. ACM Mobicom 2003*.

[8] Borst, S.C. (2003). User-level performance of channel-aware scheduling algorithms in wireless data networks. In: *Proc. Infocom 2003*.

[9] Borst, S.C., Whiting, P.A. (2003). Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization. *IEEE Trans. Veh. Techn.* **52**, 569–586.

[10] Chang, C.S., Chao, X., Pinedo, M. (1991). Monotonicity results for queues with doubly stochastic Poisson arrivals: Ross' conjecture. *Adv. Appl. Prob.* **23**, 210–228.

[11] Chaponniere, E.F., Black, P.J., Holtzman, J.M., Tse, D.N.C. (2002). Transmitter directed code division multiple access system using path diversity to equitably maximize throughput. US Patent 6,449,490.

[12] Dai, J.G. (1995). On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Prob.* **5**, 49–77.

[13] Grossglauser, M., Tse, D.N.C. (2002). Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. Netw.* **10**, 477–486.

[14] Holtzman, J.M. (2000). CDMA forward link waterfilling power control. In: *Proc. IEEE VTC 2000 Spring Conf.*, 1663–1667.

[15] Hu, T., Pan, X. (2000). Comparisons of dependence for stationary Markov processes. *Prob. Eng. Inf. Sc.* **14**, 299–315.

[16] Jalali, A., Padovani, R., Pankaj, R. (2000). Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In: *Proc. IEEE VTC 2000 Spring Conf.*, 1854–1858.

[17] Kelly, F.P. (1979). *Reversibility and Stochastic Networks*, Wiley.

[18] Kushner, H.J., Whiting, P.A. (2002). Asymptotic properties of Proportional Fair sharing algorithms. In: *Proc. 40th Annual Allerton Conf. Commun. Control., Comp.*, 1051–1059.

[19] Liu, X., Chong, E.K.P., Shroff, N.B. (2003). A framework for opportunistic scheduling in wireless networks. *Comp. Netw.* **41**, 451-474.

[20] Meyn, S.P., Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*, Springer Verlag.

[21] Müller, A., Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, Wiley.

[22] Proutière A. (2003). *Insensitivity and stochastic comparisons in queueing networks*. PhD dissertation, Ecole Polytechnique, Palaiseau, France.

[23] Shakkottai, S., Stolyar, A.L. (2001). Scheduling algorithms for a mixture of real-time and non-real time data in HDR. In: *Proc. ITC-17*, 793–804.

[24] Stolyar, A.L.. On the asymptotic optimality of the gradient scheduling algorithm for multi-user throughput allocation. To appear in *Oper. Res.*.

[25] Tchen, A.H. (1980). Inequalities for distributions with given marginals. *Ann. Appl. Prob.* **8**, 812–827.

[26] Tsibonis, V., Georgiadis, L., Tassiulas, L. (2003). Exploiting wireless channel state information for throughput maximization. In: *Proc. Infocom 2003*.