
Flow-Level Performance and User Mobility in Wireless Data Networks

Author(s): Sem Borst

Source: *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, Vol. 366, No. 1872, Networks: Modelling and Control (Jun. 13, 2008), pp. 2047-2058

Published by: [The Royal Society](#)

Stable URL: <http://www.jstor.org/stable/25190811>

Accessed: 12/06/2013 23:03

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The Royal Society is collaborating with JSTOR to digitize, preserve and extend access to *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*.

<http://www.jstor.org>

Flow-level performance and user mobility in wireless data networks

BY SEM BORST^{1,2,*}

¹*Bell Labs, Alcatel-Lucent, PO Box 636, Murray Hill, NJ 07974, USA*

²*Department of Mathematics and Computer Science, Eindhoven University
of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands*

Channel conditions in wireless networks exhibit huge variations across space and time, giving rise to vast fluctuations in the transmission rates. Channel-aware scheduling strategies provide an effective mechanism for improving throughput performance by exploiting such rate variations, and these have been extensively examined at the packet level for a static user configuration. In this paper, we discuss the performance implications at the flow level for a dynamic user population, taking into account variations on a slower time scale and wide-range user mobility as well. First of all, we present simple necessary conditions for flow-level stability and prove that these are in fact (near) sufficient for a wide family of utility-based scheduling strategies. It is further shown how the flow-level performance of the proportional fair scheduling strategy may be evaluated by means of a processor-sharing model with a state-dependent service rate. In addition, we examine the impact of variations on a slower time scale, and establish that the so-called fluid and quasi-stationary regimes yield explicit, insensitive performance bounds. Finally, we turn our attention to a network of several base stations (BSs) with handoffs of active sessions governed by wide-range user mobility. It is demonstrated that mobility increases the capacity, not only in the case of globally optimal scheduling but also when each of the BSs adheres to a local fair-sharing discipline.

Keywords: channel-aware scheduling; fluid regime; flow-level performance; proportional fair scheduling; quasi-stationary regime; user mobility

1. Introduction

The design and efficient operation of wireless networks involve similar issues to those encountered in wireline situations, such as congestion control, fairness and service differentiation. In wireless environments, these issues are greatly exacerbated by interference problems, intrinsically limited bandwidths and highly uncertain and random characteristics of wireless channels. Specifically, the channel quality may differ widely among spatially distributed users due to distance-related attenuation. In addition, the channel conditions for a given user may vary dramatically over time owing to fading effects.

* Address for correspondence: Bell Labs, Alcatel-Lucent, PO Box 636, Murray Hill, NJ 07974, USA (sem@research.bell-labs.com).

One contribution of 16 to a Discussion Meeting Issue ‘Networks: modelling and control’.

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. Multi-path fading arises at the level of a wavelength and occurs on a time scale that depends on the propagation characteristics, the carrier frequency and the speed of mobility. Path loss and shadow fading manifest themselves at a more macroscopic level as a result of distance-related attenuation and scattering due to obstacles and terrain conditions, and tend to vary over a longer time scale.

Wireless voice systems rely on power control mechanisms for adjusting the transmit power to combat fading and maintain a constant 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 an intrinsic rate requirement. Such elastic applications are well suited for rate control algorithms that continuously adapt the transmission rate over time so as to match the fluctuations in channel quality. Channel-aware or opportunistic scheduling strategies offer an effective mechanism for improving throughput performance by exploiting such rate variations (Knopp & Humblet 1995), and these have been intensively examined at the packet level for a static user configuration (e.g. Borst & Whiting 2001, 2003; Agrawal & Subramanian 2002; Neely *et al.* 2002, 2003, 2005; Viswanath *et al.* 2002; Liu *et al.* 2003; Tsibonis *et al.* 2003; Andrews *et al.* 2004, 2005; Bonald 2004; Kushner & Whiting 2004; Stolyar 2004, 2005*a,b*; Eryilmaz & Srikant 2005; Lin & Shroff 2005). We refer to the excellent survey by Georgiadis *et al.* (2006) for a comprehensive overview and additional references.

In this paper, we explore the implications of rate variations for performance at the flow level in wireless data networks, particularly borrowing from results originally reported in Borst (2003, 2005), Bonald *et al.* (2004*a,b*), Borst & Jonckheere (2006) and Borst *et al.* (2006). Besides the throughput gains produced by channel-aware scheduling, we also explicitly account for the impact of rate variations on a slower time scale and handoffs of active sessions as governed by wide-range user mobility. Owing to page limitations, we present a high-level discussion and refer to a longer version of the paper for technical details (Borst 2008).

We first present simple necessary conditions for flow-level stability and establish that these are in fact (near) sufficient for a wide family of utility-based scheduling strategies (Kelly *et al.* 1998; Mo & Walrand 2000). Imposing certain symmetry assumptions on the rate variations, we then proceed to show how the flow-level performance of the proportional fair scheduling strategy may be evaluated by means of a processor-sharing model with a state-dependent service rate. The latter model provides explicit formulae for the distribution of the number of active users, mean transfer delays, blocking probabilities and user throughputs. In particular, the performance is *insensitive*, in the sense that these measures depend only on the statistical characteristics of the system through a readily computed load factor.

Next, we focus on the impact of user mobility that occurs on a slower time scale and manifests itself in the form of rate variations at the flow level, but remains confined to a single cell. Owing to these slower rate variations, the insensitivity property is lost and the performance depends in some complicated fashion on the detailed rate statistics and traffic characteristics. In order to obtain

tractable performance estimates, we examine two limit regimes, termed *fluid* and *quasi-stationary* regimes, and use stochastic comparison techniques to show that these yield optimistic and conservative performance estimates, respectively.

Finally, we turn our attention to a network of several base stations (BSs) with handoffs of active sessions governed by wide-range user mobility. The dynamic interaction among neighbouring BSs due to interference and mobility is quite complex and has received little attention in the literature. As one of the few exceptions, Bonald *et al.* (2004*a,b*) considered a network in which the transmission rates at the various BSs depend on the activity states of surrounding BSs as a result of interference. We characterize the capacity of networks with several BSs and show that mobility increases the capacity region, not only in the case of globally optimal scheduling but also when each of the BSs adheres to a local fair-sharing discipline such as round robin.

The rest of this paper is organized as follows. In §2, we first establish simple necessary conditions for flow-level stability, and then prove that these are (near) sufficient for a wide family of utility-based scheduling strategies. In §3, we describe how the flow-level performance of the proportional fair scheduling strategy may be evaluated by means of a multi-class processor-sharing model with a state-dependent service rate. In §4, we examine the impact of rate variations on a slower time scale and establish that two limit regimes, termed fluid and quasi-stationary regimes, yield explicit, insensitive performance bounds. Finally, in §5, we turn our attention to a network of several BSs with handoffs of active sessions governed by wide-range user mobility, and show that mobility increases the capacity region. We make some concluding remarks in §6.

2. Flow-level stability

We consider a system with K traffic classes, each representing a category of statistically identical users in terms of service requirements and rate characteristics. Class- k users arrive as a Poisson process of rate λ_k (per time unit) and have generally distributed service requirements F_k (in bits) with mean ξ_k . Let $\sigma_k := \lambda_k \xi_k$ denote the offered traffic of class k (in bits per time unit) and we define $\sigma := (\sigma_1, \dots, \sigma_K)$.

For a given user population (n_1, \dots, n_K) , let $A(n_1, \dots, n_K) \subseteq \mathbb{R}_+^K$ be the set of feasible service rate vectors. The dependence on the user population accounts for the fact that the scheduling gains from fast rate variations increase with the degree of multi-user diversity (see, for instance, ch. 6 of Tse & Viswanath 2005). Observe that the sets $A(n_1, \dots, n_K)$ are convex and monotone increasing in the user population, i.e. if $(m_1, \dots, m_K) \leq (n_1, \dots, n_K)$, then $A(m_1, \dots, m_K) \subseteq A(n_1, \dots, n_K)$, since additional users may simply be excluded from service without affecting the feasible service rates of the remaining users. We define $A^* \subseteq \mathbb{R}_+^K$ as the closure of $\bigcup_{(n_1, \dots, n_K) \in \mathbb{N}^K} A(n_1, \dots, n_K)$, which inherits the convexity of the sets $A(n_1, \dots, n_K)$. While the set A^* may have a complicated structure in general, it has a rather simple form in the special case where only a single user is served in each time slot: $A^* = \{x \in \mathbb{R}_+^K : \sum_{k=1}^K x_k / R_k^* \leq 1\}$, with R_k^* denoting the maximum possible rate value of class- k users.

Proposition 2.1 provides a simple necessary stability condition.

Proposition 2.1 (Borst & Jonckheere 2006). *No scheduling strategy achieves flow-level stability for $\sigma \notin A^*$.*

Proposition 2.1 follows from the simple observation that the long-term mean service rate vector must be contained in $\text{conv}(A^*) = A^*$.

Henceforth, we focus on the family of one-parameter utility functions introduced by Mo & Walrand (2000), i.e. $U(x) = x^{1-\alpha}/(1-\alpha)$ for some $\alpha \geq 0$, with the convention that $U(x) = \log(x)$, for $\alpha = 1$. The latter family of utility functions covers the most common fairness notions, such as max-throughput ($\alpha = 0$), proportional fairness ($\alpha = 1$) and max-min fairness ($\alpha = \infty$). Note that α -fair strategies may be implemented at the slot level via (adaptive) weight-based (gradient-based) scheduling strategies (Viswanath *et al.* 2002; Stolyar 2004, 2005a,b).

Proposition 2.2 shows that the condition identified in proposition 2.1 is in fact (near) sufficient for stability of weighted α -fair scheduling strategies, assuming exponential service requirement distributions.

Proposition 2.2 (Borst & Jonckheere 2006). *A weighted α -fair scheduling strategy with $\alpha > 0$ achieves flow-level stability for $\sigma \in \text{int}(A^*)$.*

Proposition 2.2 may be established in a similar fashion as the flow-level stability of weighted α -fair rate allocation policies in wireline networks (Bonald & Massoulié 2001).

In the special case where only a single user is served in each time slot so that $A^* = \{x \in \mathbb{R}_+^K : \sum_{k=1}^K x_k/R_k^* \leq 1\}$, the stability condition of proposition 2.2 reduces to $\sum_{k=1}^K \sigma_k/R_k^* < 1$. Thus, a weighted α -fair strategy achieves stability as long as the system is stable if every class could be served at the maximum possible rate whenever selected for service. This may be explained by the observation that, under a weighted α -fair strategy, every class will either be served at the maximum possible rate or not at all whenever any of the classes is unstable.

It is interesting to observe that propositions 2.1 and 2.2 contrast with the fact that utility-based scheduling strategies generally fail to provide maximum stability guarantees at the packet level (e.g. Andrews 2004; Neely *et al.* 2005). Various simple queue-length-based strategies, on the other hand, do achieve stability at the packet level whenever possible (Tassiulas & Ephremides 1992, 1993; Andrews *et al.* 2004). In order to reconcile these paradoxical facts, it is worth observing that while such utility-based strategies operate agnostically of the queue lengths at the packet level, they *do* respond to congestion that occurs at the flow level. Thus, from a stability perspective, the behaviour of a utility-based strategy at the flow level shows resemblance to that of a queue-length-based strategy at the packet level.

3. Proportional fair scheduling

In §2, we derived simple necessary conditions for flow-level stability and established that these are in fact (near) sufficient for a wide family of α -fair scheduling strategies. We now specifically consider the proportional fair strategy (Kelly *et al.* 1998; Chaponniere *et al.* 2002), which corresponds to $\alpha = 1$ in a

scenario where the distribution of the set of feasible rate vectors is symmetric in the sense that the relative variations around the time-average rates are statistically identical for all users. In that case, each user receives service at a fraction $G(n)/n$ of its time-average rate when there are n users in the system, where $G(n)$ represents a scheduling gain factor (Borst 2008). As shown below, we can not only establish stability criteria but also explicitly evaluate the flow-level performance in terms of the number of active users, mean delays, blocking probabilities and user throughputs, and even allow for generally distributed service requirements.

As before, we consider K traffic classes and let \bar{R}_k denote the rate coefficient of a class- k user. Let $B_k := F_k/\bar{R}_k$ be the normalized service requirement of a class- k user with mean $\beta_k := \xi_k/\bar{R}_k$. We define $\rho_k := \lambda_k \beta_k = \sigma_k/\bar{R}_k$ as the normalized traffic intensity of class k and $\rho := \sum_{k=1}^K \rho_k$ as the total normalized traffic intensity. Let B_k^r be a random variable representing the residual lifetime of B_k , and $B_k^r(\cdot)$ the corresponding distribution function, i.e. $B_k^r(x) := \mathbb{P}\{B_k^r < x\} := 1/\beta_k \int_{y=0}^x \mathbb{P}\{B_k > y\} dy$. In addition, we include admission control and assume that, at most, M users are admitted in the system simultaneously. Users which initiate service requests when there are already M transfers in progress are denied access and abandoned.

Let (N_1, \dots, N_K) be a random vector representing the number of users of the various classes at an arbitrary epoch in statistical equilibrium (assuming it exists). Let $N := N_1 + \dots + N_K$ denote the total number of users in the system. Given that there are n_k class- k users in the system, let $B_{k,i}^r$ be the remaining normalized service requirement of the i th class- k user, $i=1, \dots, n_k$, $k=1, \dots, K$.

As mentioned above, each class- k user receives service at rate $G(n)\bar{R}_k/n$ when there are n active users in the system. Thus, the normalized remaining service requirement of each user is reduced at rate $G(n)/n$, which means that the normalized remaining service requirements evolve in a similar probabilistic fashion as the remaining service requirements in a multi-class processor-sharing system with arrival rates λ_k , generic service requirements B_k and service rate $G(n)$ when there are, in total, n users present. Proposition 3.1 follows from well-known results for such a system (Cohen 1979; Kelly 1979).

Proposition 3.1 (Borst 2003, 2005). *The proportional fair strategy achieves stability for $\rho < G^*$ or $M < \infty$, in which case*

$$\begin{aligned} & \mathbb{P}\{N_k = n_k, B_{k,i}^r \leq t_{k,i}; i = 1, \dots, n_k, k = 1, \dots, K\} \\ &= H^{-1} \frac{n! \rho^n}{\phi(n)} \prod_{k=1}^K \frac{1}{n_k!} \left(\frac{\rho_k}{\rho}\right)^{n_k} \prod_{i=1}^{n_k} B_k^r(t_{k,i}), \end{aligned}$$

with $n := \sum_{k=1}^K n_k \leq M$, $\phi(n) := \prod_{i=1}^n G(i)$ and normalization constant $H := \sum_{n=0}^M \rho^n / \phi(n)$.

Using Little's law, we find that the mean delay experienced by a class- k user is given by $\mathbb{E}\{S_k\} = (\beta_k/(\rho(1-L)))\mathbb{E}\{N\}$, with $L = \mathbb{P}\{N = M\}$. The above formula reflects the insensitivity property of the processor-sharing discipline, which shows that the mean delay of a class- k user depends only on the service requirement distribution of class k through its mean β_k . In fact, it may be shown

that the conditional expected delay of any user with actual service requirement b is given by $\mathbb{E}\{S|B=b\} = (b/(\rho(1-L)))\mathbb{E}\{N\}$. Thus, the expected delay incurred by a user is proportional to its normalized service requirement, with proportionality factor $\mathbb{E}\{N\}/(\rho(1-L))$.

4. Slow rate variations

So far, we have assumed that the rate variations occur on a fast time scale and average out over the time scale of interest for the flow-level performance, only manifesting themselves in the throughput gains obtained from channel-aware scheduling. We now turn our attention to a scenario where the fluctuations may have a slowly varying component. We continue to assume that the rate variations are symmetric in the sense described in §3, but now additionally allow the rate coefficients of the various users to slowly vary over time. In order to model the slower fluctuations, it is convenient to adopt a ‘state structure’, with the users moving among several possible states indexed by a finite set \mathcal{I} . The states implicitly correspond to subregions of the cell area. When in state $i \in \mathcal{I}$, the rate coefficient of a user is C_i .

As before, we consider a system with K traffic classes and now let F_{km} denote the service requirement of the m th arriving class- k user, and $S_{km}(t)$ denote its state at time t . We assume that F_{km} and $S_{km}(t)$, $m=1, 2, \dots$, are independent and identically distributed copies of a random variable F_k with mean ξ_k and a stationary and ergodic process $S_k(t)$ with state space \mathcal{I} , respectively. Let $R_k(t) = C_{S_k(t)}$ denote the generic rate process for class k and we define $\bar{R}_k := \mathbb{E}\{R_k(0)\}$ as the expected rate coefficient of a class- k user. Assuming proportional fair scheduling, the service rate at the flow level of the m th arriving class- k user, if present at time t , is $C_{S_{km}(t)} G(n)/n$, where n denotes the total number of users present at time t and $G(n)$ represents the scheduling gain factor mentioned earlier.

In order to obtain tractable performance estimates, we introduce two limit regimes, termed *fluid* and *quasi-stationary* regimes, where the rate processes evolve on an infinitely fast and slow time scale, respectively. Formally, let us consider a family of systems, parametrized by $s \in (0, \infty)$, where the generic rate process for class k is $R_k^{(s)}(t) \equiv R_k(st)$. Thus, the parameter s represents the ‘speed’ of the rate process, or, equivalently, the value $1/s$ models the time scale of the rate process. If $S_k(t)$ 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 $s \rightarrow \infty$, the variations completely vanish and the rate process reduces to a constant, giving rise to the ‘fluid’ regime with $R_k^{\text{fl}}(t) \equiv R_k^{(\infty)}(t) = \bar{R}_k$. It is worth observing that the fluid regime is reminiscent of but different from the usual law-of-large-numbers fluid limit. 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 $s \rightarrow 0$, the changes completely disappear and the rate process freezes in some initial state, yielding the ‘quasi-stationary’ regime with $R_k^{\text{qs}}(t) \equiv R_k^{(0)}(t) = R_k(0)$, where $R_k(0)$ has the stationary marginal distribution of the process $R_k(t)$.

Now observe that in the fluid and quasi-stationary regimes, the system behaves as a multi-class processor-sharing system with a state-dependent service rate $G(n)$ and class- k service requirements given by F_k/\bar{R}_k and $F_k/R_k(0)$, respectively.

Accordingly, define the traffic intensities associated with class k in the fluid and quasi-stationary regimes as $\rho_k^{\text{fl}} := \lambda_k \xi_k / \bar{R}_k$ and $\rho_k^{\text{qs}} := \lambda_k \xi_k \mathbb{E}\{1/R_k(0)\}$, respectively, with $\rho_k^{\text{fl}} \leq \rho_k^{\text{qs}}$ owing to Jensen's inequality. We denote the total traffic intensities in the fluid and quasi-stationary regimes by $\rho^{\text{fl}} := \sum_{k=1}^K \rho_k^{\text{fl}}$ and $\rho^{\text{qs}} := \sum_{k=1}^K \rho_k^{\text{qs}}$, respectively.

With the above translation to a multi-class processor-sharing system, the performance in the fluid and quasi-stationary regimes may be explicitly evaluated using well-known results for such a system (Cohen 1979; Kelly 1979). In particular, a necessary and sufficient condition for stability of the fluid (respectively, quasi-stationary) regime is $\rho^{\text{fl}} < G^*$ (respectively, $\rho^{\text{qs}} < G^*$). When the system is stable, the stationary distributions π^{fl} and π^{qs} of the numbers (n_1, \dots, n_K) of active flows of the various classes in the fluid and quasi-stationary regimes depend on the class characteristics through the traffic intensities ρ_k^{fl} and ρ_k^{qs} only,

$$\pi^{\#}(n_1, \dots, n_K) = \pi^{\#}(0) \frac{n}{\phi(n)} \prod_{k=1}^K \frac{(\rho_k^{\#})^{n_k}}{n_k!},$$

where ‘#’ represents ‘fl’ or ‘qs’; $n := \sum_{k=1}^K n_k$; $\phi(n) := \prod_{i=1}^n G(i)$; and $\pi^{\#}(0)$ is determined by the normalizing condition.

It follows from the stochastic bounds to be derived below that the condition $\rho^{\text{fl}} < G^*$ is necessary for stability, while the condition $\rho^{\text{qs}} < G^*$ is sufficient. Note that when the number of active flows tends to infinity, each flow stays in the system for a long time, so that the rate process tends to average out over the flow duration, i.e. the system behaves as in the fluid regime. Thus, one would expect the condition $\rho^{\text{fl}} < G^*$ to be not only necessary but also sufficient for stability, as confirmed by proposition 4.1.

Proposition 4.1 (Bonald et al. 2004b). *If $\rho^{\text{fl}} < G^*$, then the system is stable.*

We now compare the performance of the system with that in the fluid and quasi-stationary regimes. Assume that the system is empty at time 0 and let $N_k(t)$ denote the number of active class- k flows at time t . With minor abuse of notation, for $m=1, \dots, N_k(t)$, let $\hat{F}_{km}(t)$ be the remaining service requirement of the m th active class- k flow at time t . We define the total workload at time t as $W(t) := \sum_{k=1}^K \sum_{m=1}^{N_k(t)} \hat{F}_{km}(t) / \bar{R}_k$. We assume that the cumulative distribution function $P(\cdot) = \mathbb{P}\{F \leq \cdot\}$ associated with the service requirement F is concave.

Proposition 4.2 states that the performance improves (respectively, deteriorates) in terms of the number of active flows, the workload and the response times D , when the rate processes of some flows are replaced by the corresponding fluid (respectively, quasi-stationary) versions as described earlier. Similar comparison results hold for the equilibrium distributions.

Proposition 4.2 (Bonald et al. 2004b). *For all $k=1, \dots, K$, $t \geq 0$,*

$$W^{\text{fl}}(t) \leq_{icx} W(t) \leq_{icx} W^{\text{qs}}(t), \quad N_k^{\text{fl}}(t) \leq_{st} N_k(t) \leq_{st} N_k^{\text{qs}}(t), \quad D^{\text{fl}} \leq_{st} D \leq_{st} D^{\text{qs}},$$

where \leq_{icx} and \leq_{st} indicate the increasing-convex and stochastic orderings, respectively (Müller & Stoyan 2002), while the superscript ‘fl’ (respectively, ‘qs’) refers to a system where the rate processes of some flows are replaced by the corresponding fluid (respectively, quasi-stationary) versions.

5. Large-scale user mobility

We now further extend the model to a network with several BSs, indexed by the set \mathcal{N} , and possible handoffs among cells as governed by large-scale user mobility in addition to slow rate variations driven by mobility within cell areas. The scheduling gains from fast rate variations are no longer included. In order to model the slow rate variations and possible handoffs, we continue to use the state structure introduced before, with users moving among several possible states. However, we now additionally allow the users to be served by different BSs in different states. Define \mathcal{I}_n as the set of states where a user is served by BS n .

We consider a system with K traffic classes as described in §4. As before, the state of a class- k user is governed by a stationary ergodic process $S_k(t)$, with $\pi_{i,k} := \mathbb{P}\{S_k(t) = i\}$. Let $\mathcal{J}_k := \{i \in \mathcal{I} : \pi_{i,k} > 0\}$ denote the set of states in which a class- k user may reside. We define $p_{k,n} := \sum_{i \in \mathcal{I}_n} \pi_{i,k}$ as the stationary probability that a class- k user is in cell n and let $\mathcal{N}_k := \{n \in \mathcal{N} : p_{k,n} > 0\}$ denote the set of cells that class- k users may visit.

(a) Optimal scheduling strategies

We now determine the capacity region for optimal scheduling strategies. In other words, we characterize the set of all traffic vectors $(\sigma_1, \dots, \sigma_K)$ for which there exists a scheduling strategy that achieves stability. We define

$$\mathcal{T} := \left\{ (\tau_{i,k})_{i \in \mathcal{I}, k=1, \dots, K} \in \mathbb{R}_+^{(|\mathcal{I}| \times K)} : \tau_{i,k} = 0 \text{ if } i \notin \mathcal{J}_k, \text{ and } \sum_{k=1}^K \sum_{i \in \mathcal{I}_n \cap \mathcal{J}_k} \tau_{i,k} = 1, \forall n \in \mathcal{N} \right\},$$
$$\mathcal{R} := \left\{ (r_1, \dots, r_K) \in \mathbb{R}_+^K : \exists \tau \in \mathcal{T} \text{ such that } r_k \leq \sum_{i \in \mathcal{I}} \tau_{i,k} C_i, \forall k = 1, \dots, K \right\}.$$

The component $\tau_{i,k}$, $i \in \mathcal{I}_n$, may be interpreted as the fraction of resources of BS n allocated to class- k users in state i . With that interpretation, the quantity $\sum_{i \in \mathcal{I}} \tau_{i,k} C_i$ represents the total service rate received by class- k users. Thus, \mathcal{R} may be interpreted as the achievable rate region, i.e. the set of all achievable service rates for the various traffic classes. Note that \mathcal{R} is a convex set and depends on the spatial user distributions through the sets \mathcal{J}_k only.

Proposition 5.1 Borst et al. (2006). *There exists a scheduling strategy that achieves stability if $(\sigma_1, \dots, \sigma_K) \in \text{int}(\mathcal{R})$. Conversely, if $(\sigma_1, \dots, \sigma_K) \notin \mathcal{R}$, then there exists no scheduling strategy that achieves stability.*

(i) No mobility

We now consider a scenario without any mobility, i.e. the state of a user is assumed to be fixed for the duration of the flow, and the probability that a class- k user is in state i is $\psi_{i,k}$. The rate region simplifies to $\mathcal{R}^{\text{no}} := \{(r_1, \dots, r_K) \in \mathbb{R}_+^K : \exists \tau \in \mathcal{T}, \text{ such that } \psi_{i,k} r_k < \tau_{i,k} C_i, \forall i \in \mathcal{I}, k = 1, \dots, K, \text{ and may equivalently be represented as } \mathcal{R}^{\text{no}} = \{(r_1, \dots, r_K) \in \mathbb{R}_+^K : \sum_{k=1}^K \sum_{i \in \mathcal{I}_n} \psi_{i,k} r_k / C_i < 1, \forall n \in \mathcal{N}\}$.

It is easily verified that $\mathcal{R}^{\text{no}} \subseteq \mathcal{R}$, assuming $\{i \in \mathcal{I} : \psi_{i,k} > 0\} \subseteq \{i \in \mathcal{I} : \pi_{i,k} > 0\}$. This demonstrates that mobility increases the achievable capacity as long as the scheduling discipline is left flexible, which agrees with findings in the context of

ad hoc mobile networks (Grossglauser & Tse 2002). However, this does not imply that mobility increases the capacity achieved by a given scheduling discipline, e.g. fair resource sharing. In this regard, it is worth recalling that the results in §4 show that *intra-cell* mobility, which manifests itself in the form of slow rate variations, improves the performance under fair resource sharing. In §5b, we investigate whether *inter-cell* mobility increases the capacity under fair resource sharing as well.

(b) *Fair resource sharing*

We now turn our attention to a scenario where each of the BSs operates according to an α -fair scheduling strategy. Thus, when there are n_{ik} class- k users in state i , $i \in \mathcal{I}_n$, each of them receives service at rate

$$\frac{C_i^{1/\alpha}}{\sum_{l=1}^K \sum_{j \in \mathcal{I}_n} n_{j,l} C_j^{1/\alpha-1}}.$$

We define

$$\mathcal{R}^{\text{fair}} := \left\{ r \in \mathbb{R}_+^K : \exists \theta \in \mathbb{R}_+^K \text{ such that } r_k \leq \sum_{n \in \mathcal{N}} \frac{\sum_{i \in \mathcal{I}_n} \pi_{i,k} \theta_k C_i^{1/\alpha}}{\sum_{l=1}^K \sum_{j \in \mathcal{I}_n} \pi_{j,l} \theta_l C_j^{1/\alpha-1}}, \forall k \right\}.$$

Note that the vector $(\tau_{i,k})$, with

$$\tau_{i,k} = \frac{\pi_{i,k} \theta_k C_i^{1/\alpha-1}}{\sum_{l=1}^K \sum_{j \in \mathcal{I}_n} \pi_{j,l} \theta_l C_j^{1/\alpha-1}}, \quad i \in \mathcal{I}_n,$$

belongs to the set \mathcal{T} , so that $\mathcal{R}^{\text{fair}} \subseteq \mathcal{R}$. The components of the vector $(\theta_1, \dots, \theta_K)$ may be interpreted as the numbers of flows of the various classes. With that interpretation, the quantity

$$\sum_{n \in \mathcal{N}} \frac{\sum_{i \in \mathcal{I}_n} \pi_{i,k} \theta_k C_i^{1/\alpha}}{\sum_{l=1}^K \sum_{j \in \mathcal{I}_n} \pi_{j,l} \theta_l C_j^{1/\alpha-1}}$$

represents the total service rate received by class- k flows under an α -fair sharing strategy.

Proposition 5.2 provides a characterization of the capacity region in the case of α -fair resource sharing, assuming exponential service requirement distributions. It states that the system is stable if and only if there exists a relative distribution of the numbers of flows across the various classes so that the total service rate received by each of the traffic classes is larger than the traffic intensity of that class.

Proposition 5.2 (Borst *et al.* 2006). *If $(\sigma_1, \dots, \sigma_K) \in \text{int}(\mathcal{R}^{\text{fair}})$, then the system is stable. If $(\sigma_1, \dots, \sigma_K) \notin \mathcal{R}^{\text{fair}}$, then the system is unstable.*

(i) *No mobility*

We now consider a scenario without any mobility, i.e. the state of a user is assumed to be fixed for the duration of the flow, and the probability that a class- k user is in state i is $\psi_{i,k}$. As mentioned earlier, the rate region simplifies to $\mathcal{R}^{\text{no}} = \{(r_1, \dots, r_K) \in \mathbb{R}_+^K : \sum_{k=1}^K \sum_{i \in \mathcal{I}_n} \psi_{i,k} r_k / C_i < 1, \forall n \in \mathcal{N}\}$.

The latter region may or may not be larger than the stability region in the presence of mobility, depending on the probabilities $\pi_{i,k}$ and $\psi_{i,k}$ and the rates C_i . When the spatial user distribution is identical in the two scenarios, however, i.e. $\pi_{i,k} = \psi_{i,k}$ for all $i \in \mathcal{I}$, $k = 1, \dots, K$, then it may be shown that $\mathcal{R}^{\text{no}} \subseteq \mathcal{R}^{\text{fair}}$.

6. Conclusion

In summary, we observe that rate variations and mobility tend to improve the performance and increase the capacity region. In some respects, this simply ties in with the generic principle of opportunistic scheduling, and confirms the natural expectation that the throughput improvements obtained at the packet level translate into performance and capacity gains at the flow level. This applies not only to fast rate variations but also to mobility that occurs on a slower time scale and creates additional opportunities to serve users in the most favourable locations, which may be interpreted as a form of network-wide opportunistic scheduling. In this regard, the above observation resonates with the fact that mobility increases the capacity of ad hoc mobile networks (Grossglauser & Tse 2002). What is perhaps less obvious is that the performance and capacity improvements also arise in the case of fair sharing, e.g. channel-oblivious round-robin scheduling, where BSs allocate equal resource shares to users in less advantageous conditions. In that case, the improvements reflect a certain convexity property of flow-level performance measures and capacity with respect to the rate processes, as inspection of the proofs reveals. The above results have important implications for traffic engineering and dimensioning. In particular, ‘neglecting’ mobility and pretending that users remain stationary for the duration of their service yields a simple, conservative capacity estimate.

The author gratefully acknowledges the tremendous benefits enjoyed from fruitful interactions with Thomas Bonald (France Telecom R&D), Nidhi Hegde (France Telecom R&D), Matthieu Jonckheere (CWI), Sindo Núñez-Queija (CWI), Alexandre Proutière (Microsoft Research) and Phil Whiting (Alcatel-Lucent).

References

- Agrawal, R. & Subramanian, V. 2002 Optimality of certain channel-aware scheduling policies. In *Proc. 40th Annual Allerton Conf. Commun., Control, Comp., Monticello, IL, 2–4 October 2002*, pp. 1532–1541.
- Andrews, D. M. 2004 Instability of the proportional fair scheduling algorithm for HDR. *IEEE Trans. Wireless Commun.* **3**, 1422–1426. (doi:10.1109/TWC.2004.833419)
- Andrews, D. M., Kumaran, K., Ramanan, K., Stolyar, A. L., Vijayakumar, R. & Whiting, P. A. 2004 Scheduling in a queueing system with asynchronously varying service rates. *Prob. Eng. Inf. Sci.* **18**, 191–217. (doi:10.1017/S0269964804182041)

Phil. Trans. R. Soc. A (2008)

- Andrews, D. M., Qian, L. & Stolyar, A. L. 2005 Optimal utility-based throughput allocation subject to throughput constraints. In *Proc. Infocom 2005, Miami, FL, 15–17 March 2005*. Piscataway, NJ: IEEE Press.
- Bonald, T. 2004 A score-based opportunistic scheduler for fading radio channels. In *Proc. European Wireless Conference 2004, Barcelona, Spain, 24–27 February 2004*. Berlin, Germany: VDE Verlag.
- Bonald, T. & Massoulié, L. 2001 Impact of fairness on Internet performance. In *Proc. ACM Sigmetrics/Performance 2001, Boston, MA, 18–20 June 2001*, pp. 82–91. New York, NY: ACM Press.
- Bonald, T., Borst, S. C., Hegde, N. & Proutière, A. 2004a Wireless data performance in multi-cell scenarios. In *Proc. ACM Sigmetrics/Performance 2004, New York, NY, 14–16 June 2004*, pp. 378–388. New York, NY: ACM Press.
- Bonald, T., Borst, S. C. & Proutière, A. 2004b How mobility impacts the flow-level performance of wireless data systems. In *Proc. IEEE Infocom 2004, Hong Kong, People's Republic of China, 9–11 March, 2004*. Piscataway, NJ: IEEE Press.
- Borst, S. C. 2003 User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proc. Infocom 2003, San Francisco, CA, 1–3 April 2003*. Piscataway, NJ: IEEE Press.
- Borst, S. C. 2005 User-level performance of channel-aware scheduling algorithms in wireless data networks. *IEEE/ACM Trans. Netw.* **13**, 636–647. (doi:10.1109/TNET.2005.850215)
- Borst, S. C. 2008 Flow-level performance and user mobility in wireless data networks. Technical Report. See <http://plan9.bell-labs.com/who/sem/publications/rs220108.pdf>.
- Borst, S. C. & Jonckheere, M. 2006 Flow-level stability of channel-aware scheduling-algorithms. In *Proc. WiOpt '06, Boston, MA, 4–6 April 2006*. Piscataway, NJ: IEEE Press.
- Borst, S. C. & Whiting, P. A. 2001 Dynamic rate control algorithms for HDR throughput optimization. In *Proc. Infocom 2001, Anchorage, AK, 24–26 April 2001*, pp. 976–985. Piscataway, NJ: IEEE Press.
- Borst, S. C. & Whiting, P. A. 2003 Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization. *IEEE Trans. Vehicular Technol.* **52**, 569–586. (doi:10.1109/TVT.2003.810967)
- Borst, S. C., Hegde, N. & Proutière, A. 2006 Capacity of wireless networks with intra- and inter-cell mobility. In *Proc. Infocom 2006, Barcelona, Spain, 25–27 April 2006*. Piscataway, NJ: IEEE Press.
- 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.
- Cohen, J. W. 1979 The multiple phase service network with generalized processor sharing. *Acta Informatica* **12**, 245–284. (doi:10.1007/BF00264581)
- Eryilmaz, A. & Srikant, R. 2005 Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. In *Proc. Infocom 2005, Miami, FL, 15–17 March 2005*. Piscataway, NJ: IEEE Press.
- Georgiadis, L., Neely, M. J. & Tassiulas, L. 2006 Resource allocation and cross-layer control in wireless networks. *Found. Trends Netw.* **1**, 1–144. (doi:10.1561/13000000001)
- Grossglauser, M. & Tse, D. N. C. 2002 Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. Netw.* **10**, 477–486. (doi:10.1109/TNET.2002.801403)
- Kelly, F. P. 1979 *Reversibility and stochastic networks*. Chichester, UK: Wiley.
- Kelly, F. P., Maulloo, A. & Tan, D. 1998 Rate control for communication networks: shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.* **49**, 237–252. (doi:10.1057/palgrave.jors.2600523)
- Knopp, R. & Humblet, P. A. 1995 Information theory and power control in single-cell multi-user communications. In *Proc. ICC '95, Seattle, WA, 19–22 June 1995*. Piscataway, NJ: IEEE Press.
- Kushner, H. J. & Whiting, P. A. 2004 Convergence of proportional-fair sharing algorithms under general conditions. *IEEE Trans. Wireless Commun.* **3**, 1250–1259. (doi:10.1109/TWC.2004.830826)

- Lin, X. & Shroff, N. B. 2005 The impact of imperfect scheduling on cross-layer rate control in wireless networks. In *Proc. Infocom 2005, Miami, FL, 15–17 March 2005*. Piscataway, NJ: IEEE Press.
- Liu, X., Chong, E. K. P. & Shroff, N. B. 2003 A framework for opportunistic scheduling in wireless networks. *Comp. Netw.* **41**, 451–474. (doi:10.1016/S1389-1286(02)00401-2)
- Mo, J. & Walrand, J. C. 2000 Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.* **8**, 556–567. (doi:10.1109/90.879343)
- Müller, A. & Stoyan, D. 2002 *Comparison methods for stochastic models and risks*. New York, NY: Wiley.
- Neely, M. J., Modiano, E. & Rohrs, C. E. 2002 Power and server allocation in a multi-beam satellite with time-varying channels. In *Proc. Infocom 2002, New York, NY, 25–27 June 2002*, pp. 1451–1460. Piscataway, NJ: IEEE Press.
- Neely, M. J., Modiano, E. & Rohrs, C. E. 2003 Dynamic power allocation and routing for time-varying wireless networks. In *Proc. Infocom 2003, San Francisco, CA, 1–3 April 2003*. Piscataway, NJ: IEEE Press.
- Neely, M. J., Modiano, E. & Li, C.-P. 2005 Fairness and optimal stochastic control for heterogeneous networks. In *Proc. Infocom 2005, Miami, FL, 15–17 March 2005*. Piscataway, NJ: IEEE Press.
- Stolyar, A. L. 2004 MaxWeight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic. *Ann. Appl. Prob.* **14**, 1–53. (doi:10.1214/aoap/1075828046)
- Stolyar, A. L. 2005a Maximizing queueing network utility subject to stability: greedy primal dual algorithm. *Queueing Syst.* **50**, 401–457. (doi:10.1007/s11134-005-1450-0)
- Stolyar, A. L. 2005b On the asymptotic optimality of the gradient scheduling algorithm for multi-user throughput allocation. *Oper. Res.* **53**, 12–25. (doi:10.1287/opre.1040.0156)
- Tassiulas, L. & Ephremides, A. 1992 Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Automat. Contr.* **37**, 1936–1948. (doi:10.1109/9.182479)
- Tassiulas, L. & Ephremides, A. 1993 Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Trans. Inf. Theory* **30**, 466–478. (doi:10.1109/18.212277)
- Tse, D. N. C. & Viswanath, P. 2005 *Fundamentals of wireless communication*. Cambridge, UK: Cambridge University Press.
- Tsibonis, V., Georgiadis, L. & Tassiulas, L. 2003 Exploiting wireless channel state information for throughput maximization. In *Proc. Infocom 2003, San Francisco, CA, 1–3 April 2003*. Piscataway, NJ: IEEE Press.
- Viswanath, P., Tse, D. N. C. & Laroia, R. 2002 Opportunistic beamforming using dumb antennas. *IEEE Trans. Inf. Theory* **48**, 1277–1294. (doi:10.1109/TIT.2002.1003822)