RESEARCH   **Open Access**

Opportunistic scheduling policies for improved throughput guarantees in wireless networks

Jawad Rasool[1*], Vegard Hassel[2], Sébastien de la Kethulle de Ryhove[3] and Geir E Øien[1]

Abstract

Offering throughput guarantees for cellular wireless networks, carrying real-time traffic, is of interest to both the network operators and the customers. In this article, we formulate an optimization problem which aims at maximizing the throughput that can be guaranteed to the mobile users. By building on results obtained by Borst and Whiting and by assuming that the distributions of the users' carrier-to-noise ratios are known, we find the solution to this problem for users with different channel quality distributions, for both the scenario where all the users have the same throughput guarantees, and the scenario where all the users have different throughput guarantees. Based on these solutions, we also propose two simple and low complexity adaptive scheduling algorithms that perform significantly better than other well-known scheduling algorithms. We further develop an expression for the approximate throughput guarantee violation probability for users in time-slotted networks with the given cumulants of the distribution of bit-rate in a time-slot, and a given distribution for the number of time-slots allocated within a time-window.

1 Introduction

In modern wireless networks, *opportunistic multiuser scheduling* has been implemented to obtain a more efficient utilization of the scarcely available radio spectrum. For wireless cellular standards, such as 1 × EVDO, HSDPA, and Mobile WiMAX [1], the scheduling algorithms are often not specified in the standardization documents. The scheduling algorithms implemented might therefore vary from vendor to vendor. Selecting the most efficient scheduling algorithms will be critical for having the most efficient utilization of a wireless network; consequently, the vendors that implement the most-suited scheduling algorithms will have a competitive advantage.

Opportunistic multiuser scheduling will give higher throughput in a wireless cell than non-opportunistic algorithms like Round Robin (RR) because priority is given to the users with the most favorable channel conditions [2,3]. However, always selecting the users with the best channel quality may lead to starvation of other users.

Consequently, the quality-of-service (QoS) demands of the users also have to be taken into account when designing practical wireless scheduling algorithms. A common approach to obtain higher QoS in the network is to have a fairer resource allocation among the users [4,5]. One widely adopted fair scheduling policy is the Proportional Fair Scheduling (PFS) algorithm [6]. When there are many users in a cell, this algorithm ensures both that the users are scheduled close to their own peak carrier-to-noise ratio (CNR) and that they have the same probability of being scheduled in a randomly picked time-slot [7].

With real-time traffic transmitted over wireless networks, the need for more exact QoS measures is in the interests of both network operators and customers. The customers want to know what they have bought, and the operators would rather not give away more network capacity to the customers than they have paid for. A measure that is well suited to quantify QoS guarantees exactly is a *throughput guarantee*, i.e., how many bits a user is guaranteed to transmit or receive within a time-window. Throughput guarantees can in principle be either *hard* or *deterministic*, and *soft* or *statistical*. Hard throughput guarantees promise with unit probability that a guarantee will be fulfilled, while the corresponding soft throughput guarantees promise with a lower

* Correspondence: jawad.rasool@iet.ntnu.no
[1]Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim NO-7491, Norway
Full list of author information is available at the end of the article

than unity–but preferably high–probability that the specified throughput guarantee will be fulfilled. For telecommunications networks in general, and for wireless networks in particular, soft throughput guarantees are more suitable for specifying QoS than hard throughput guarantees. This is because such networks often have a varying number of users and varying loads from the applications of these users. For wireless networks, the varying quality of the radio channel will further add uncertainty to the size of the throughput that can be guaranteed during short time-spans.

A general framework for opportunistic scheduling is presented in [8], along with three general categories of scheduling problems under this framework. The third category, i.e., *minimum performance requirement* discusses the scenario that is similar to the proposed one in this study. A stochastic-approximation-based algorithm is also provided to estimate the key parameters of the scheduling scheme online. However, the merit and novelty of our study is that our scheduling algorithm is significantly simpler and thus more applicable than the one proposed in [8]. In addition, we show the performance in real-life networks.

In [9], Andrews et al. propose scheduling algorithms that aim at fulfilling throughput guarantees by giving different priorities to the users depending on how far they are from their maximum and minimum throughput guarantees. One of the problems with this algorithm is that it takes action only when a throughput guarantee has been violated. Andrews et al. have therefore shown in [9] how time parameters of their algorithm can be set shorter than the actual time-window of interest to alleviate this issue. In this article we propose an alternative scheduling algorithm that tries to fulfill the throughput guarantees before they are violated.

A utility-based predictive scheduler is proposed in [10] that focuses on fulfilling the throughput guarantees by predicting the future channel conditions and adopting the rates accordingly. At the current time slot, it schedules the user whose future channel conditions would make it more difficult to provide the throughput guarantees.

Borst and Whiting have elegantly proved that a certain scheduling policy provides the highest throughput guarantee for wireless networks [11]. However, they briefly argue that the rate distributions of the users are unknown, and they have therefore not shown how this optimal scheduling policy can be found for users with differently distributed CNRs. They have also not designed algorithms that will give the lowest short-term *throughput guarantee violation probability* (TGVP), which we define as the probability of not fulfilling a throughput guarantee within a specified time-window, averaged over all the users in the system. In this study,

we argue that, for many scenarios, the CNR distributions of the users can in fact be estimated, and that, we hence can use these distributions to develop efficient scheduling algorithms for providing short-term throughput guarantees.

This article collects, unifies, and discusses in depth the results in conference papers [12-14], providing a complete overview of the modeling, analysis methods, and simulation results which are only partially covered in those papers. We formulate an optimization problem aimed at finding an optimal scheduling algorithm that obtains maximum throughput guarantees in a wireless network. By building on the results in [11] and by assuming that the distributions of the users' CNRs are known, we show how the solution to this optimization problem can be obtained numerically both when the throughput guarantees are (i) the same and (ii) different for all the mobile users. We also propose two adaptive algorithms that improve the performance of the optimal algorithm for short time-windows. In real systems, some of the users are static users, while others are pedestrian or vehicular users. We therefore also analyze the performance of these algorithms for different time-slot correlations corresponding to different users' speeds. Quantifying the soft throughput guarantees for a certain scheduling algorithm, without conducting experimental investigations, is valuable for network providers. We also develop an expression for the approximate TGVP for users in time-slotted networks, for any scheduling algorithm with the given cumulants of the distribution of bit-rate in a time-slot, and a given distribution for the number of time-slots allocated within a time-window. Through simulations, we show that our TGVP approximation is tight for a realistic network, with fast moving users with correlated channels and realistic throughput guarantees.

Our proposed scheduling algorithms aim not only at fulfilling the throughput guarantees that are promised to the mobile users in a wireless network, but our analysis can also be used to estimate the expected TGVP of all the users if a new user is admitted into the system. Such real-time TGVP estimates can be useful when performing admission control.

It should be noted that our analysis involves several idealistic assumptions (see Section 2). For example, we assume that the CNR can be estimated perfectly and fed back with infinite precision and no delay, that ideal adaptive modulation and coding can be performed, that the CNR distributions of the users can be estimated perfectly, and that the population of backlogged users is constant over the time-window over which the throughput guarantees are calculated. How realistic these assumptions are for real-life networks is a subject for further research.

The rest of this article is organized as follows. In Section 2, we present the system model, and in Section 3, we formulate the optimization problem for obtaining the highest possible throughput guarantee over a time-window. In Section 4, we show how the solution to this problem can be found when all the users have the same throughput guarantees. The corresponding solution for heterogeneous throughput guarantees is discussed in Section 5. In Section 6, we derive an approximate expression for the TGVP, while we describe the novel adaptive scheduling algorithms in Section 7. In Section 8, we discuss some practical considerations before presenting our numerical results in Section 9. Section 10 focuses on related work on short-term throughput guarantees. We list our conclusions in Section 11.

2 System model

We consider a single base station that serves N-backlogged users using time-division multiplexing (TDM). The analysis conducted in this article is valid both for the uplink and the downlink; in either case we assume that the total available bandwidth for the users is W [Hz] and that the users have constant transmit power. Each user estimates his own CNR perfectly, and before performing downlink scheduling, the base station is assumed to receive these measurements from all the users. The base station also performs uplink scheduling based on perfect channel estimates, and for each time-slot, the base station takes a scheduling decision and distributes this decision to the selected user before uplink transmission starts.

It is assumed that the communication channel between the base station and the users can be modeled by a flat, block-fading channel, subject to additive white Gaussian noise; moreover, that the communication channels corresponding to the different users fade independently. The block duration equals one time-slot and is denoted T_{TS} [seconds]. We also assume that the CNR values corresponding to different time-slots are correlated. The correlation model used in our simulations will be described in detail in Section 8.

The average CNR of user i is denoted by $\bar{\gamma}_i$. Without loss of generality, we assume that the user indices are assigned in a manner such that user 1 has the lowest average CNR, user 2 has the second lowest average CNR, and so on, down to user N, which has the highest average CNR. Assuming constant average CNR values for the time-window over which the throughput guarantees are calculated can be realistic for a real-life wireless network. This is because the average CNR of the users' CNR distributions normally changes on a time-scale of several seconds while the throughput guarantees are often calculated over time-windows of less than one hundred milliseconds.

We also assume that the probability distributions of the CNRs of each of the users are perfectly known (however, a known *joint* CNR distribution is *not* required). In modern cellular standards like 1 × EVDO, HSDPA, and Mobile WiMAX [1], much of the information needed for obtaining precise probability distribution estimates is already available. To conduct adaptive coding and modulation, modern cellular networks have precise, real-time CNR estimates of the users. These channel quality estimates can therefore be utilized to obtain estimates of the probability distributions of the CNRs of each one of the users. Such probability distribution estimates can be obtained from some hundred CNR estimates by using, e.g., order statistic filter banks [15]. To further improve the estimates of the probability distributions, we can adapt the estimation techniques to the types of terrain that the users operate in and to the speed of the users. For example, for a channel with many reflectors, with no line-of-sight (LOS) component, and with a relatively high speed of the users, a Rayleigh channel model will give a good estimate of the distribution of the channel gain. When we have a LOS component, a Rice channel can be assumed.

Another important assumption is that the population of backlogged users is constant and equal to N. According to [11], this assumption is realistic since the separation of time-scales makes the population of backlogged users nearly static; i.e., the population of backlogged users changes much slower than the time-window over which the throughput guarantees are calculated.

3 The optimization problem

We now formulate an optimization problem aimed at obtaining the maximal throughput guarantee B [bits], which can be achieved within a time-window of T_W [seconds]. A similar optimization problem has also been formulated in [11] and explored in [12,13]. In this section, we assume that the same throughput guarantee is promised to all the users, i.e.,

$$T_i \bar{R}_i = B, \qquad (1)$$

for all $i = 1, ..., N$, where T_i [seconds] is the accumulated time allocated to user i over the time-window and $\bar{R}_i$ [bits/s] is the average rate for user i when he/she is transmitting or receiving. By virtue of the TDM assumption, the sum of the T_i's satisfies

$$\sum_{i=1}^{N} T_i = T_W. \qquad (2)$$

Under the assumption that T_i is long enough to make the time-window T_W infinitely long, (1) can also be written as

$$p(i)T_W \bar{R}_i = B, \tag{3}$$

where $p(i)$ is the access probability for user i within the time-window T_W. From (1)-(3), we obtain

$$p(i) = \frac{1}{\bar{R}_i \sum_{j=1}^{N} \frac{1}{\bar{R}_j}}. \tag{4}$$

Assuming that T_i is long enough and contains enough time-slots for the channel to reveal its ergodic properties, and that the Shannon capacity can be achieved, the average rate $\bar{R}_i$ for user i when he/she is transmitting or receiving, can be written as

$$\bar{R}_i = W \int_0^\infty \log_2(1 + \gamma) p_{\gamma*}(\gamma|i) d\gamma, \tag{5}$$

where $p_{\gamma*}(\gamma|i)$ is the probability density function (PDF) of the CNR of user i when this user is scheduled. From the equations above, our objective is to find a *scheduling policy* that gives the maximum B that can be promised to all the users over the time-window T_W, meaning that (1) has to be maximized subject to the constraints (5), for $i = 1, ..., N$. We show in the next section how to obtain this optimal scheduling policy.

4 Solution to the optimization problem

It was shown in [11] that the following scheduling algorithm gives the solution to the optimization problem described in the previous section:

$$i^*(t_k) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\frac{r_i(t_k)}{\alpha_i} \right), \tag{6}$$

where $i^*(t_k)$ is the index of the user that is going to be scheduled in time-slot k, $r_i(t_k)$ is the instantaneous rate of user i in time-slot k, and α_i is a constant. However, in [11], it is not shown how the optimal α_i's can be found. If we assume that the PDFs of the users' channel gains are known, and that we have an ideal link adaptation protocol and block-fading, then we can use this result to obtain a solution to the optimization problem in the previous section. To obtain this solution, we define the random variable $S_i \triangleq \frac{R_i}{\alpha_i}$, where R_i is the random variable describing the rate of user i. S_i is the scheduling metric of the algorithm, i.e., the metric that decides which user is going to be scheduled. For flat, block-fading channels, the maximal value of the metric S_i for user i within a time-slot (block) with CNR γ can be expressed as

$$S_i(\gamma) = \frac{W \log_2(1 + \gamma)}{\alpha_i}. \tag{7}$$

In real-life systems, we can come close to this maximum value of S_i by using efficient link adaptation and (close-to-)capacity-achieving codes. Assuming Rayleigh faded channel gains, and denoting by $p_{\gamma_i}(\gamma)$ the PDF of the CNR of user i, the PDF for the normalized rate $S_i = s$ for user i can be written as

$$p_{S_i}(s) = \left. \frac{p_{\gamma_i}(\gamma)}{\frac{dS_i(\gamma)}{d\gamma}} \right|_{\gamma = 2^{\frac{s \cdot \alpha_i}{W}} - 1} = \frac{\alpha_i \ln(2)}{W \bar{\gamma}_i} 2^{\frac{s \cdot \alpha_i}{W}} e^{-\frac{2^{\frac{s \cdot \alpha_i}{W}} - 1}{\bar{\gamma}_i}}. \tag{8}$$

The corresponding cumulative distribution function can be expressed as

$$P_{S_i}(s) = \int_0^s p_{S_i}(x) dx = 1 - e^{-\frac{2^{\frac{s \cdot \alpha_i}{W}} - 1}{\bar{\gamma}_i}}. \tag{9}$$

We can now express the access probability of user i as

$$p(i) = \int_0^\infty p_{S_i}(s) \prod_{\substack{j=1 \\ j \neq i}}^{N} P_{S_j}(s) ds. \tag{10}$$

Furthermore, the PDF of S_i when user i is scheduled can be found by using *Bayes' rule*:

$$p_{S_i}(s|i) = \frac{p_{S_i}(s)}{p(i)} \prod_{\substack{j=1 \\ j \neq i}}^{N} P_{S_j}(s). \tag{11}$$

We can also express the expected value of S_i conditioned on user i being scheduled, as

$$\mathrm{E}[S_i|i] = \frac{\mathrm{E}[R_i|i]}{\alpha_i} = \frac{\bar{R}_i}{\alpha_i} = \int_0^\infty s p_{S_i}(s|i) ds. \tag{12}$$

Combining (4), (10), and (12) we obtain $3N$ equations in $3N$ unknowns, and can thus find the values for the $p(i)$'s, the $\bar{R}_i$'s, and the α_i's. A solution can be found by using numerical integration together with an algorithm for solving sets of nonlinear equations. This can, for example, be achieved in *MATLAB* by using the functions `quad` and `fsolve`. It should be noted that it has not been proved that the solution to this set of equations is unique. Note that when $\alpha_i = \alpha$ for all $i = 1, ..., N$, the scheduling algorithm given in (6) reduces to Maximum CNR Scheduling (MCS) algorithm, which schedules the user with the highest CNR, and hence the highest rate.

Since this scheduling algorithm maximizes B, we would expect that this algorithm will yield higher values

of B than any of the other classical scheduling algorithms. However, one should remember that it is implicitly assumed in (1) that the average rate of the users over the time-window equals their expected throughput. This will only be true when the time-window T_W can be considered infinitely long and contains infinitely many time-slots. The solution is consequently suboptimal for short time-windows containing only a small amount of time-slots. In Section 7, we therefore propose two adaptive scheduling algorithms that show good performance also for short time-windows with few time-slots.

5 Optimization for heterogeneous throughput guarantees

When the throughput guarantees are different from user to user, we can again use the scheduling policy corresponding to (6), but with a different set of α_i's to obtain the optimal bit allocation. By using B_i [bits] to denote the throughput guarantee for user i during the time-window T_W, we obtain

$$T_i \bar{R}_i = B_i. \tag{13}$$

Equation 2 becomes

$$\sum_{i=1}^{N} \frac{B_i}{\bar{R}_i} = T_W. \tag{14}$$

For a finite but long time-window T_W, we have[a]

$$p(i) T_W \bar{R}_i \approx B_i. \tag{15}$$

From (14) and (15), we obtain the following expression for $p(i)$:

$$p(i) \approx \frac{B_i}{\bar{R}_i \sum_{j=1}^{N} \frac{B_j}{\bar{R}_j}}. \tag{16}$$

We can now fix the throughput guarantees B_i of up to N - 1 users and maximize the remaining throughput guarantees by solving the set of $3N$ equations resulting from (16), (10), and (12). To be able to solve this optimization problem, we can, for example, additionally constrain the users with non-fixed B_i's to have equal throughput guarantees. It is also important to note that setting fixed throughput guarantees that are too high will yield an optimization problem with no solution–meaning that such throughput guarantees are not achievable by the system. Of course, it only makes sense to set fixed throughput guarantees that are achievable by the system.

6 Throughput guarantee violation probability
The TGVP is defined as the probability of not fulfilling a throughput guarantee B [bits] within a specified time-window T_W [seconds], averaged over all N users in the system [16]. For a specific user i, the TGVP_i is the probability of the number of bits b_i transmitted to or from it within a time-window T_W being below B_i, and is denoted as

$$\mathrm{TGVP}_i = \mathrm{Pr}(b_i < B_i), \quad i = 1, 2, \ldots, N. \tag{17}$$

In this study, we focus on the TGVP because a throughput guarantee in most cases cannot be given with absolute certainty, i.e., we are focusing on *soft* throughput guarantees. The guaranteed number of bits B_i within the time-window T_W should, however, be promised to the users with high probability. This means that when assessing the relative behavior of different scheduling algorithms, the TGVP performance of the algorithms close to TGVP = 0 is the most interesting.

6.1 Deriving (approximate) TGVP expression
In this subsection, we derive an expression for TGVP that can be used as a tool to specify an achievable soft throughput guarantee of B bits over a time-window T_W constituting K time-slots, for users transmitting over a time-slotted block fading channel.

In [16], an approximate expression for TGVP is also derived by using the central limit theorem. Although that expression provides a very good TGVP approximation, we argue that since the users are generally offered (soft) throughput guarantees with close to unit probability, the probability of violating a throughput guarantee should be very small, i.e., close to zero. In this derivation, we therefore argue that a non-zero TGVP should be treated as a *rare* event. Large deviation theory (LDT) is a branch of probability theory that deals with rare events and provides asymptotic estimates for their probabilities. We shall use Cramer's theorem [17, p. 27] from LDT to derive the approximate TGVP expression in what follows. (This approach was initially proposed by us in [14].)

The allocation of different numbers of time-slots to a user constitutes mutually exclusive events. The TGVP for user i over K time-slots can therefore be expressed as follows, using the law of total probability:

$$\begin{aligned}
\mathrm{Pr}(b_i < B) \quad = \quad & \mathrm{Pr}(b_i < B|0) \cdot p_M(0|i) \\
& + \mathrm{Pr}(b_i < B|1) \cdot p_M(1|i) \\
& \cdots \\
& + \mathrm{Pr}(b_i < B|K) \cdot p_M(K|i),
\end{aligned} \tag{18}$$

where $\mathrm{Pr}(b_i < B|k)$ denotes the TGVP when user i is assigned $M = k$ time-slots, and $p_M(k|i)$ denotes the probability that user i gets $M = k$ time-slots within the interval of K time-slots.

To be able to discuss a total throughput guarantee B within K time-slots, we first consider the number of bits transmitted to or from user i within the jth time-slot

he/she is scheduled, and denote this number by $b_{i,j}$, with $\mu_{b_{i,j}}$ and $\sigma^2_{b_{i,j}}$ the mean and variance of $b_{i,j}$, respectively. For a system using constant transmit power and capacity-achieving codes which operate at the Shannon capacity limit, we will have $b_{i,j} = T_{\text{TS}}W\log_2(1 + \gamma_{i,j})$, where $\gamma_{i,j}$ is the CNR in the jth time-slot user i is scheduled.

We can now express the probability for violating the throughput guarantee B when k out of K time-slots are scheduled to user i as

$$Pr(b_i < B|k) = \Pr\left(\sum_{j=1}^{k} b_{i,j} < B\right)$$
$$= \Pr\left(\bar{b}_{i,k} < \frac{B}{k}\right), \quad (19)$$

where $\bar{b}_{i,k} = \frac{1}{k}\sum_{j=1}^{k} b_{i,j}$ is the average number of bits being transmitted to or from user i when he/she is allocated $M = k$ time-slots, and we assume that $\mu_{\bar{b}_{i,k}}$ and $\sigma^2_{\bar{b}_{i,k}}$ are the mean and variance of $\bar{b}_{i,k}$, respectively.

Next we apply Cramer's theorem by considering the following two cases:

For $\frac{B}{k} < \mu_{b_{i,j}}$, we have

$$\lim_{k\to\infty} \frac{1}{k}\log\Pr\left(\bar{b}_{i,k} \le \frac{B}{k}\right) = -I\left(\frac{B}{k}\right)$$
$$\Rightarrow \Pr\left(\bar{b}_{i,k} < \frac{B}{k}\right) \approx e^{-kI(B/k)}, \quad (20)$$

and for $\frac{B}{k} > \mu_{b_{i,j}}$,

$$\lim_{k\to\infty} \frac{1}{k}\log\Pr\left(\bar{b}_{i,k} \ge \frac{B}{k}\right) = -I\left(\frac{B}{k}\right)$$
$$\Rightarrow \Pr\left(\bar{b}_{i,k} > \frac{B}{k}\right) \approx e^{-kI(B/k)}, \quad (21)$$
$$\Rightarrow \Pr\left(\bar{b}_{i,k} < \frac{B}{k}\right) \approx 1 - e^{-kI(B/k)},$$

where $I(\cdot)$ is known as the *large deviation rate function* [17, p. 28]. It is defined as the Legendre-Fenchel transform [18] of the cumulant generating function $\lambda(\theta)$:

$$I\left(\frac{B}{k}\right) \triangleq \sup_\theta \left(\theta\frac{B}{k} - \lambda(\theta)\right). \quad (22)$$

The cumulant generating function $\lambda(\theta)$ is the logarithm of the moment generating function $M(\theta)$, and its Taylor expansion is given as follows:

$$\lambda(\theta) = \log M(\theta) = \kappa_1\theta + \kappa_2\frac{\theta^2}{2!} + \kappa_3\frac{\theta^3}{3!} + \cdots$$

The cumulants κ_1, κ_2, κ_3, ... can be calculated from the moments of the distribution of $b_{i,j}$ as follows:

$$\kappa_1 = m_1 = \mu_{b_{i,j}},$$
$$\kappa_2 = m_2 - m_1^2 = \sigma^2_{b_{i,j}},$$
$$\kappa_3 = m_3 - 3m_2m_1 + 2m_1^3,$$
$$\vdots$$

where m_l is the lth order moment of the distribution of $b_{i,j}$.

In this study, we only consider the first two cumulants for simplification. However, we must emphasize that higher order cumulants should be used for more accurate results. The cumulant generating function is then given as

$$\lambda(\theta) = \theta\mu_{b_{i,j}} + \frac{\sigma^2_{b_{i,j}}}{2}\theta^2. \quad (23)$$

Substituting (23) in (22),

$$I\left(\frac{B}{k}\right) = \sup_\theta \left(\theta\frac{B}{k} - \theta\mu_{b_{i,j}} - \frac{\sigma^2_{b_{i,j}}}{2}\theta^2\right). \quad (24)$$

The value of θ^* that maximizes (24) is found to be

$$\theta^* = \frac{\frac{B}{k} - \mu_{b_{i,j}}}{\sigma^2_{b_{i,j}}}. \quad (25)$$

Thus, the rate-function in this case is given as

$$I\left(\frac{B}{k}\right) = \frac{\left(\frac{B}{k} - \mu_{b_{i,j}}\right)^2}{2\sigma^2_{b_{i,j}}}. \quad (26)$$

Finally, the probability that the throughput constraint B is violated over K time-slots for user i can be approximated as

$$Pr(b_i < B) \approx p_M(0|i) + \sum_{k=1}^{K} p_M(k|i)\Pr(b_i < B|k), \quad (27)$$

where $Pr(b_i < B|k)$ is given in (20) and (21) for the two cases discussed.

The TGVP for the overall system is then given as

$$\text{TGVP} = \frac{1}{N}\sum_{i=1}^{N} Pr(b_i < B). \quad (28)$$

6.2 TGVP for the optimal scheduling algorithm
In this section, we focus on the optimal scheduling algorithm described in Section 4, and derive the equations

for $p_M(k|i)$, $\mu_{b_{i,j}}$ and $\sigma^2_{b_{i,j}}$ used in the TGVP expression.

All the users in our system model have the same distribution for their relative CNRs [19] and the *relatively best user* (i.e., the user with the highest $r_i(t_k)=\alpha_i$) is scheduled in each time-slot. Therefore, the number of time-slots allocated to a user i within K time-slots is distributed according to the *binomial distribution* [20, p. 1179]:

$$p_M(k|i) = \binom{K}{k} p(i)^k (1 - p(i))^{K-k},\qquad(29)$$

where $p(i)$ is the access probability of user i given in (10).

The mean $\mu_{b_{i,j}}$ and the variance $\sigma^2_{b_{i,j}}$ of $b_{i,j}$ are given as follows:

$$\mu_{b_{i,j}} = m_1 = \mathrm{E}[b_{i,j}],\qquad(30)$$

$$\sigma^2_{b_{i,j}} = m_2 - m_1^2 = \mathrm{E}[b^2_{i,j}] - (\mathrm{E}[b_{i,j}])^2.\qquad(31)$$

The first moment m_1 (the mean value for $b_{i,j}$) for our optimal scheduling algorithm is derived as follows:

$$\begin{aligned}m_1 = \mathrm{E}[b_{i,j}] &= WT_{\mathrm{TS}} \int_0^\infty \log_2(1+\gamma) p_{\gamma*}(\gamma|i) d\gamma \\ &= \alpha_i T_{\mathrm{TS}} \int_0^\infty s p_{S_i}(s|i) ds.\end{aligned}\qquad(32)$$

Using (12),

$$\mathrm{E}[b_{i,j}] = T_{\mathrm{TS}} \bar{R}_i.\qquad(33)$$

Similarly, the second moment m_2 of the number of bits $b_{i,j}$ transmitted to or from user i can be obtained as follows:

$$\begin{aligned}m_2 = \mathrm{E}[b^2_{i,j}] &= (WT_{\mathrm{TS}})^2 \int_0^\infty (\log_2(1+\gamma))^2 p_{\gamma*}(\gamma|i) d\gamma \\ &= (\alpha_i T_{\mathrm{TS}})^2 \int_0^\infty s^2 p_{S_i}(s|i) ds.\end{aligned}\qquad(34)$$

Through simulations (see Section 9), we shall show that our TGVP approximation is tight for a realistic network with fast moving users and correlated channels.

7 Adapting weights to increase short-term performance

As already mentioned, the scheduling algorithms obtained in the previous sections are only efficient when the throughput guarantees are promised over a long time-window T_W containing many time-slots. To fulfill throughput guarantees for shorter time-windows with fewer time-slots, we propose two adaptive scheduling schemes in this section.

7.1 Adaptive scheduling algorithm 1

The values of α_i found in the previous sections aim at providing throughput guarantees within *any* time-window T_W. This means that these parameters are optimized in a manner which is such that the throughput guarantees should be fulfilled independently of the time instants at which T_W starts or ends. In this subsection, we instead develop an algorithm that will only aim at fulfilling the throughput guarantees within the duration of a fixed time-window T_W. To improve performance for shorter time-windows with fewer time-slots, it is useful to *adapt* the values of the parameters α_i to the *actual resource allocation that has already been done* within the finite time-window T_W. This adaptation can be optimally done during each time slot by using the approach of the previous section with B_i / T_W replaced by $B'_i / T'_W = (B_i - B_{ik})/(T_W - T_k)$, where B_{ik} is the number of bits assigned to user i after k time-slots within the time-window T_W, and $T_k = kT_{\mathrm{TS}}$. The adaptation of the parameters α_i should in many cases be performed in time intervals of less than a millisecond. Since it can be difficult to conduct the optimal optimization described above in such a short time, we propose the following simple adaptive scheduling algorithm as an alternative:

$$i^*(t_k) = \operatorname*{argmax}_{1 \le i \le N} \left(\rho_i(t_{k-1}) \frac{r_i(t_k)}{\alpha_i} \right),\qquad(35)$$

where $\rho_i(t_k)$ is the ratio

$$\rho_i(t_k) = \frac{\max(0, B_i - B_{ik})}{T_W - T_k} \frac{T_W}{B_i}.\qquad(36)$$

The rationale behind this scheduling algorithm is as follows: The value of $\rho_i(t_k)$ expresses the normalized share of the throughput guarantee that is to be fulfilled in the remaining K - k time-slots of the time-window T_W. If the rate guarantee is already fulfilled, then the value of $\rho_i(t_k)$ is zero, which means that the user in question is not selected in the remaining K - k time-slots. If a user has been allocated exactly $\frac{B_i T_k}{T_W}$ bits after k time-slots, then the value of $\rho_i(t_k)$ will be unity, which means that this user will be scheduled with the same weights as for the non-adaptive policy. For the case where the number of allocated bits after k time slots is lower than $\frac{B_i T_k}{T_W}$ bits, the value of $\rho_i(t_k)$ will be above unity, which means that the user is given higher priority compared to the non-adaptive optimal scheduling policy. Likewise, a user is given lower priority if

he/she has been allocated more than $\frac{B_i T_k}{T_W}$ bits after k time-slots. The priority is determined by the urgency of fulfilling the throughput guarantee within the remainder of the time-window.

A similar strategy has been employed in [21] for improving short-term throughput of utility-based scheduling in CDMA wireless networks.

The problem with the above algorithm is that it can only fulfill the throughput guarantees when the placement of the window is fixed. That is, for every new time-window, the algorithm starts over again and tries to achieve the throughput guarantees. This means that the throughput guarantees cannot be promised within time-windows with a different duration or a different placement than that used by the algorithm. The consequence of this approach is that we may have to adjust the time-window T_W to the bit-streams from different speech and video codecs.

7.2 Adaptive scheduling algorithm 2

In this subsection, we describe another adaptive scheduling algorithm[b] that overcomes the problem of fixed window placement of Algorithm 1. Furthermore, this algorithm is also simpler in implementation. This novel adaptive scheduling algorithm works as follows:

For promised throughput guarantees B_i, select a user $i^*(t_k)$ that has a maximum

$$i^*(t_k) = \operatorname*{argmax}_{1 \leq i \leq N} \left(\upsilon_i(t_{k-1}) \frac{r_i(t_k)}{\alpha_i} \right), \tag{37}$$

where $\upsilon_i(t_k)$ is given as

$$\upsilon_i(t_k) = \begin{cases} 0 & \text{if } B_{ik} \geq B_i, \\ 1 & \text{otherwise}, \end{cases} \tag{38}$$

where B_{ik} is the total number of bits assigned to user i during k time slots.

The rationale behind this scheduling algorithm is very simple: If the throughput guarantee of user i is already fulfilled, then it is not selected in the remaining time-slots, i.e., the value of $\upsilon_i(t_k)$ is set to zero. For all the other users, $\upsilon_i(t_k) = 1$ so that among them, a user j is selected with maximum $r_j(t_k)/\alpha_j$.

Note that this adaptive algorithm is independent of the duration and placement of the time-window T_W.

We can intuitively say that the offline parameter α_i increases the throughput fairness of the system, whereas the online parameters ρ_i and υ_i improve the corresponding short-term performance of the system.

8 Practical considerations

In this section, we briefly discuss some practical issues as well as realistic system parameters. Interested readers are referred to [12] for a detailed discussion.

Different classes of traffic will need different values for B_i. For example, B_i/T_W can vary between 5 and 64 kbit/s for a one-way telephony speech connection [22]. For a real-life network, we can assume that the B_i's correspond to the sum of all the throughput guarantees promised to the different real-time sessions of a user. Hence, for each new video conferencing or speech connection, the network has to update the B_i's and do the optimization of the scheduling algorithm all over again.

For the wireless standards HSDPA and Mobile WiMAX, the time-slot length for the downlink is 2 and 5 ms, respectively [1]. The European IST research project WINNER I has suggested a time-slot duration of 0.34 ms for a future wireless system [23]. The corresponding time-slot length for the 3GPP LTE network is 1 ms [24]. If we assume that $T_W = 80$ ms, then the time-window contains 235, 80, 40, and 16 time-slots for WINNER I, LTE, HSDPA, and Mobile WiMAX, respectively.

If the average CNR of one or more users change or the CNR distribution of one or more users change, e.g., from Rayleigh to Rice, then the whole optimization problem has to be solved again to obtain new values for the α_i's, which is a feasible task. It should be noted that the adaptive factors $\rho_i(t_k)$ and $\upsilon_i(t_k)$ are independent of the CNR distributions.

It is more difficult to fulfill throughput guarantees for all the users in a system that has strongly temporally correlated channels, since one user can be allocated many consecutive time-slots. The temporal correlation of the channel is both dependent on the speed v of the users and on the carrier frequency f_c of the channel. For the simulations in the next sections, we assume Jakes' correlation model. The channel gain can in this case be modeled as a sum of sinusoids correlated according to $f_D T_{TS}$, where $f_D = \dfrac{v f_c}{c}$ is the Doppler frequency shift, and c is the speed of light [25].

9 Numerical results
9.1 Identical throughput guarantees

In this section, we consider the case where all the users are promised identical throughput guarantees $B = T_W$, where $T_W = 80$ ms. Figures 1, 2, 3, 4 show the TGVP performance in networks that are, respectively, based on Mobile WiMAX, HSDPA, LTE, and WINNER I. For these plots, we have assumed that only one user can be scheduled in a time-slot. As mentioned earlier, we focus on the TGVP here since a throughput guarantee in most cases cannot be given with absolute certainty. Also, the TGVP performance of the algorithms close to TGVP = 0 is the most interesting. The results are shown for 10 users having Rayleigh fading channels with average CNRs given in Table 1. The total average CNR
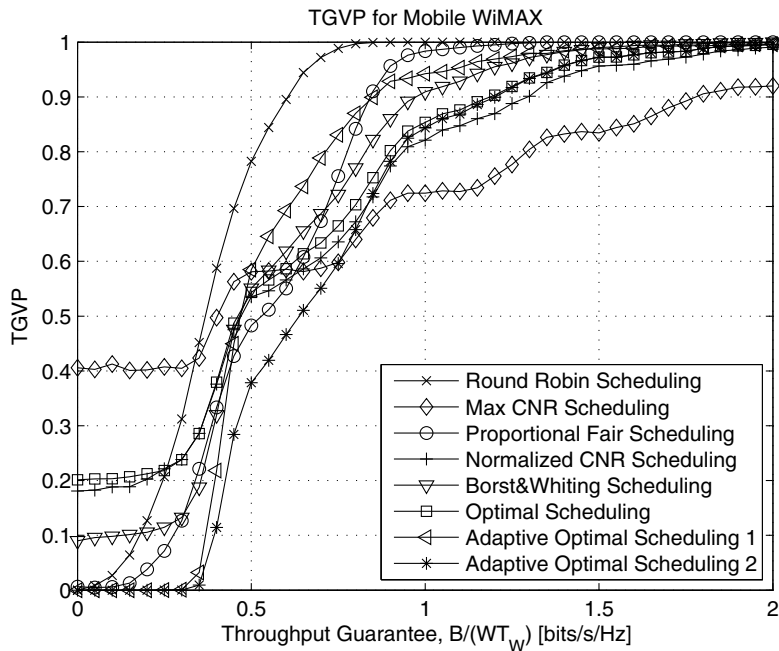
Figure 1 Throughput guarantee violation probability for 10 users in a Mobile WiMAX network with identical throughput guarantees.
Plotted for a time-window T_W = 80 ms that contains 16 time-slots. Each value in the plot is an average over 1,000 Monte Carlo simulations.
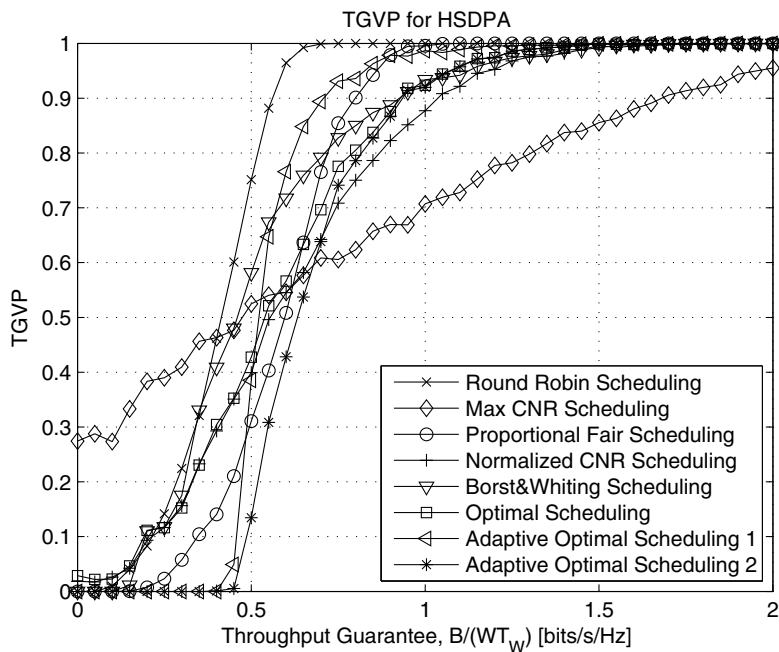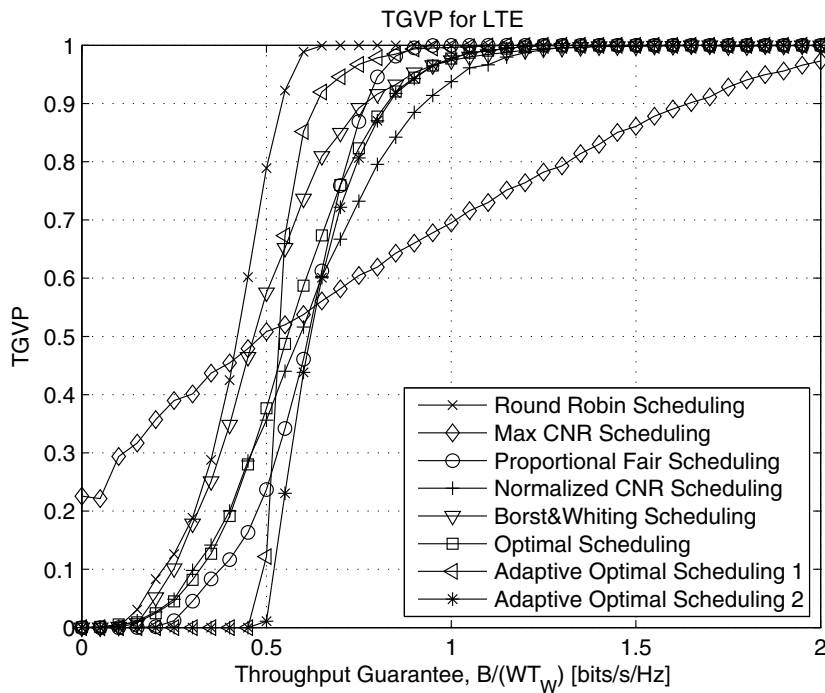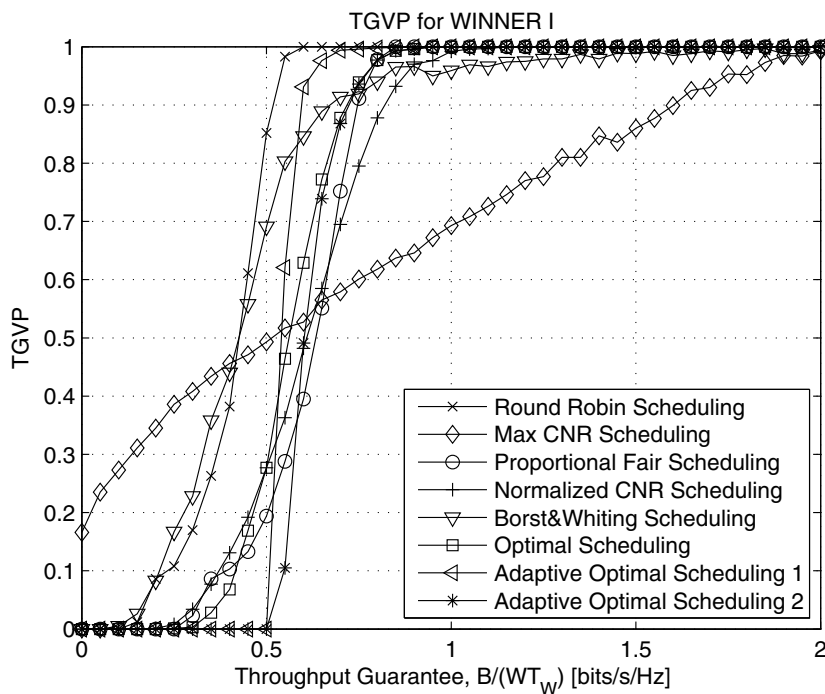

Figure 2 Throughput guarantee violation probability for 10 users in a HSDPA network with identical throughput guarantees. Plotted
for a time-window T_W = 80 ms that contains 40 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

Figure 3 Throughput guarantee violation probability for 10 users in an LTE network with identical throughput guarantees. Plotted for a time-window T_W = 80 ms that contains 80 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

Figure 4 Throughput guarantee violation probability for 10 users in a WINNER I network with identical throughput guarantees. Plotted for a time-window T_W = 80 ms that contains 235 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

Table 1 Example of parameters for 10 Rayleigh-distributed users, with identical throughput guarantees

i	$\bar{\gamma}_i$ (dB)	$\bar{\gamma}_i$	$p(i)$	$\bar{R}_i$ (bit/s)	α_i
1	5.0000	3.1623	0.180356	3.146640	2.751868
2	9.7712	9.4868	0.120250	4.719458	4.510686
3	11.9897	15.8114	0.104241	5.444237	5.354891
4	13.4510	22.1359	0.095904	5.917567	5.916338
5	14.5424	28.4605	0.090533	6.268622	6.337987
6	15.4139	34.7851	0.086660	6.548775	6.675953
7	16.1394	41.1096	0.083694	6.780829	6.958115
8	16.7609	47.4342	0.081318	6.978923	7.200381
9	17.3045	53.7587	0.079353	7.151803	7.412701
10	17.7875	60.0833	0.077691	7.304769	7.601701

is 15 dB (averaged over all the users). From (3), we can see that $B_{\mathrm{opt}}/(WT_W) = p(i)\bar{R}_i/W$ for the optimized values of $p(i)$ and $\bar{R}_i$. It is easily seen by using the values in Table 1 to calculate the product $p(i)\bar{R}_i$ for $1 \le i \le 10$ that $B_{\mathrm{opt}}/(WT_W) = 0.5675$ bits/s/Hz for all the users for this particular example. The correlation between the different time-slot CNRs has been described by Jakes' model with $f_c = 1$ GHz and a user speed of $v = 30$ m/s. It should be noted that this correlation will be stronger for short time-slots than for long time-slots. We compare the new scheduling policies to five other algorithms, namely, RR Scheduling, MCS, Normalized CNR Scheduling (NCS), PFS, and the *adaptive* scheduling algorithm proposed by Borst and Whiting in [11]. For the RR policy, the time-slots are allocated to the users in a sequential manner, i.e., totally non-opportunistically. The most opportunistic algorithm is the MCS since it always schedules the user with the highest CNR. The NCS policy is a fairer policy because it schedules the users with the highest CNR-to-average-CNR ratio. A similar policy, the PFS algorithm, schedules the user with the highest instantaneous rate divided by a weighted sum of the rate allocated in the previous time-slots [3]. For our simulations, we have implemented the PFS algorithm as described in [3], with the time-constant $t_c = T_W$ and with the initial average rate for each user equal to the theoretical average rate for this user. The adaptive Borst and Whiting algorithm is implemented as described in [11, p. 575] with $\delta(k) = 0.5 * 0.9^k$, where k denotes the kth "reset." The "price updates" of this algorithm are done every $10 * n$th time-slot, where n denotes the nth "price update." To investigate the performance of the adaptive updating of the weights for this algorithm, we have used the optimal weights as initial weights.

Figures 1, 2, 3, 4 show the TGVP as a function of $B/(WT_W)$ for a time-window of, respectively, 16, 40, 80, and 235 time-slots. We see that our novel adaptive

algorithms perform better than all the other algorithms for all the cases. It should also be noted that since the WINNER I system has many time-slots within the time-window of 80 ms, the two adaptive algorithms obtain a throughput guarantee that is very close to the optimal throughput guarantee of 0.5675 bits/s/Hz for this system. It is also interesting to observe that the throughput guarantee that can be promised with close to unity probability with the adaptive algorithms is more than twice as large as for the PFS algorithm for all the four systems. It should be noted that our non-adaptive optimal algorithm also performs better than all the other well-known algorithms for the case where the time-window contains 235 time-slots (WINNER I). The reason for this is that the non-adaptive algorithm is designed for long time-windows containing many time-slots.

9.2 Heterogeneous throughput guarantees
Analyzing the TGVP for a network where the users have heterogeneous throughput guarantees requires a significant number of plots. The analysis in this section is based on an LTE-based system, and we again consider 10 users having Rayleigh fading channels with average CNRs given in Table 1.

To simplify the analysis, we shall fix the throughput guarantees B_i/WT_W of four users to the same value, and maximize the throughput guarantees for the remaining users by solving the set of $3N$ equations resulting from (16), (10), and (12). We further constrain the users with non-fixed B_i's to have equal throughput guarantees. The two sets of four users with fixed throughput guarantees, which we use in this section, are $\{1, 2, 3, 4\}$ and $\{7, 8, 9, 10\}$. These sets correspond to the users with low and high average CNRs, respectively.

We first fix $B_i/WT_W = 0.3$ bits/s/Hz for the four users and try to maximize the throughput guarantee B/WT_W for the remaining six users. Note that these fixed throughput guarantees are lower than the $B_{\mathrm{opt}}/WT_W = 0.5675$ bits/s/Hz of the case of identical throughput guarantees. From Figures 5 and 6, we observe that the two adaptive optimal scheduling algorithms outperform all the other algorithms. The performance of the non-adaptive scheme is worse as compared to the scenario with identical throughput guarantees. Since this scheme is designed for a long time-window, it assumes that users that require 0.3 bits/s/Hz will get it in the long run. Therefore the selection of α_i's is such that lesser weight is given to these users, to maximize the throughput of the remaining users. The performance of the adaptive algorithms is better than the previous case because four users require a throughput guarantee that is less than $B_{\mathrm{opt}}/WT_W = 0.5675$ bits/s/Hz of the
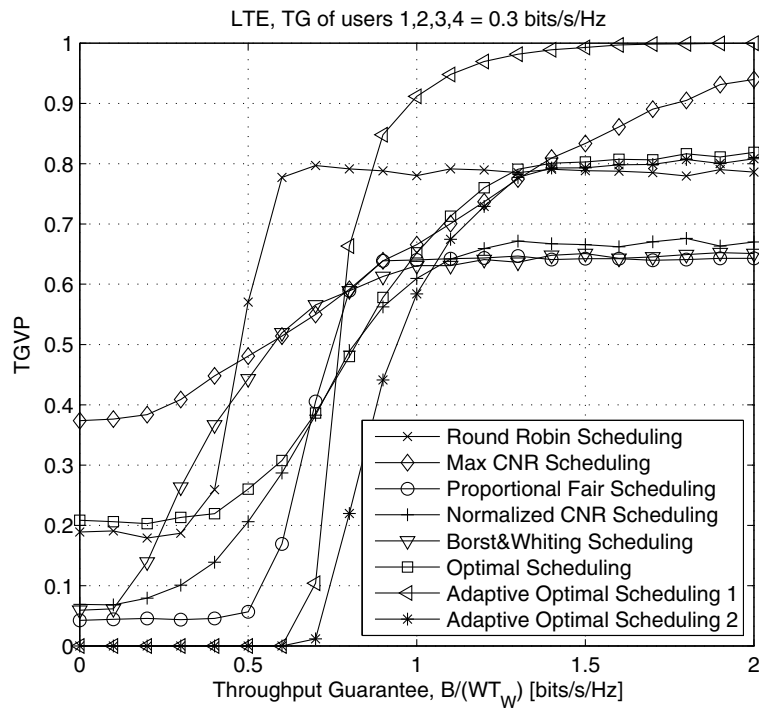
Figure 5 Throughput guarantee violation probability for 10 users in a LTE network. Throughput guarantees of users 1,2,3,4 are fixed to 0.3 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.
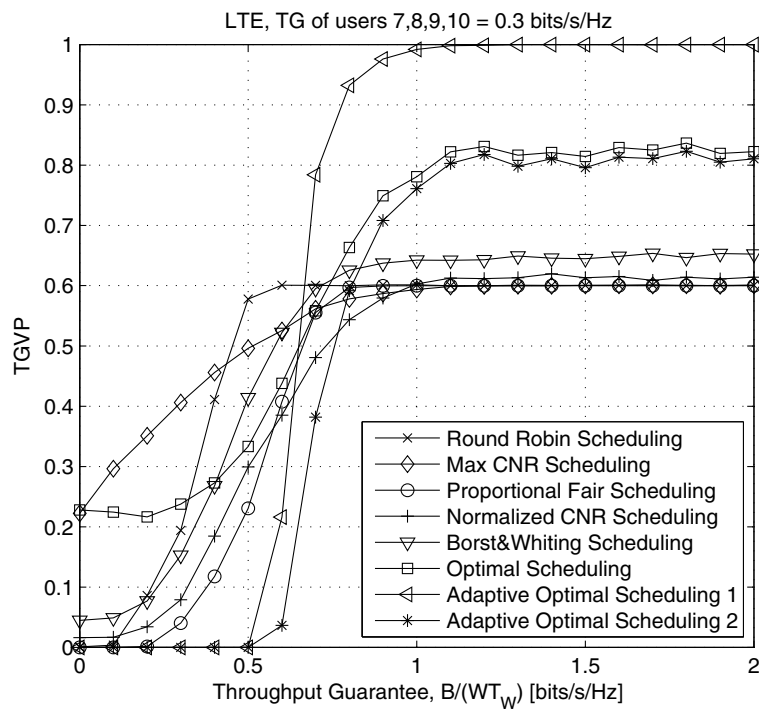

Figure 6 Throughput guarantee violation probability for 10 users in a LTE network. Throughput guarantees of users 7,8,9,10 are fixed to 0.3 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.

identical throughput guarantees case. The system allocates these extra bits to the remaining six users.

Next, we fix $B_i/WT_W = 0.7$ bits/s/Hz for the four users. This case is shown in Figures 7 and 8. Since we have selected fixed throughput guarantees higher than the $B_{\mathrm{opt}}/WT_W = 0.5675$ bits/s/Hz of the identical throughput guarantees case, the performance of all the algorithms is bound to suffer. However, it is interesting to observe that our novel adaptive optimal scheduling algorithms are still able to provide close to unity throughput guarantee, i.e., TGVP $\cong 0$. In fact, they are the only algorithms that are able to do so in this case.

9.3 Effect of temporal correlation on TGVP
One user can be allocated many consecutive time-slots if the CNR of the users are correlated from time-slot to time-slot. It is therefore more difficult to fulfill throughput guarantees for all the users in a system that has strongly temporally correlated channels. As mentioned before, the temporal correlation of the channel is both dependent on the speed v of the users and on the carrier frequency f_c of the channel. For fixed f_c, higher user speed means higher Doppler frequency shift f_D, which results in rapidly changing user channel and lower correlation between the time-slots. Thus, we expect a better TGVP performance at higher speeds. In this section, we observe (Figure 9) the

effect of users' speed on the TGVP performance of different algorithms by considering a Mobile WiMAX network with identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz. We again use Jakes' correlation model with $f_c = 1$ GHz.

As expected, the TGVP performance of RR scheduler remains the same for various users' speeds since the users are selected irrespective of their CNRs. The PFS algorithm does not suffer much at lower speeds since it takes into account the average throughput of all the users. The performance of MCS and the optimal scheduling algorithm at lower speeds deteriorate because they are not able to allocate enough time-slots to all the users due to strong temporal correlation. The performance of the adaptive algorithms at lower users' speeds remains the best. It is because these scheduling algorithms ignore the users who have already received their share, and provide sufficient time-slots to other users to fulfill their throughput guarantees within the remaining time-window.

9.4 Accuracy of the TGVP expression
Figures 10 and 11 give comparisons of the approximate TGVP expression for the optimal algorithm with the corresponding Monte Carlo-simulated TGVPs for Mobile WiMAX- and LTE-based networks, respectively, for identical throughput guarantees. The approximate results are based on the assumption that the time-slots
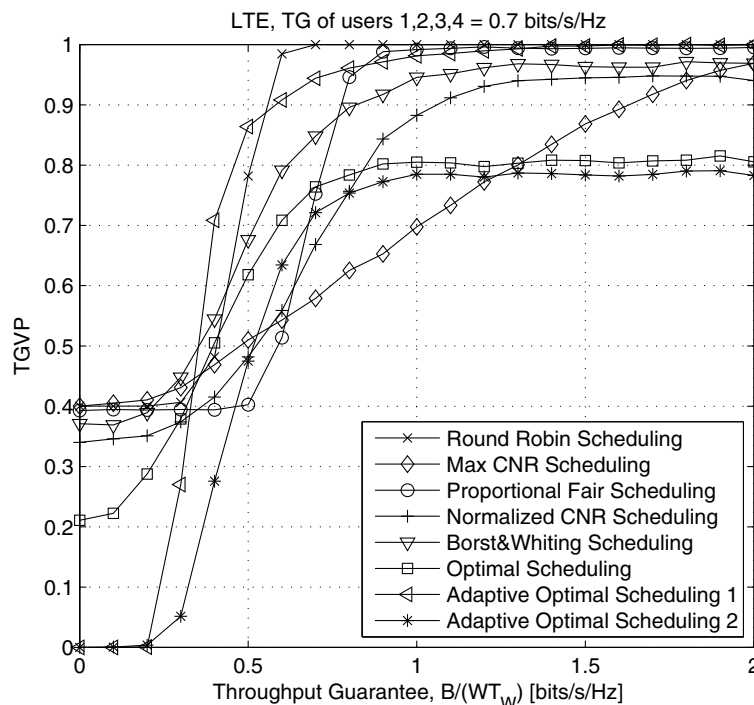

Figure 7 Throughput guarantee violation probability for 10 users in a LTE network. Throughput guarantees of users 1,2,3,4 are fixed to 0.7 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.
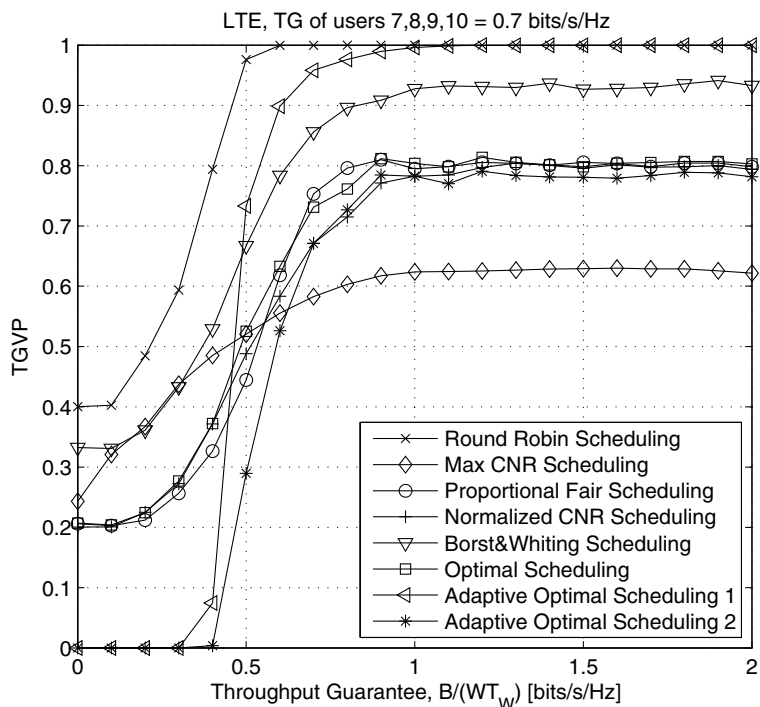
Figure 8 Throughput guarantee violation probability for 10 users in a LTE network. Throughput guarantees of users 7,8,9,10 are fixed to 0.7 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.
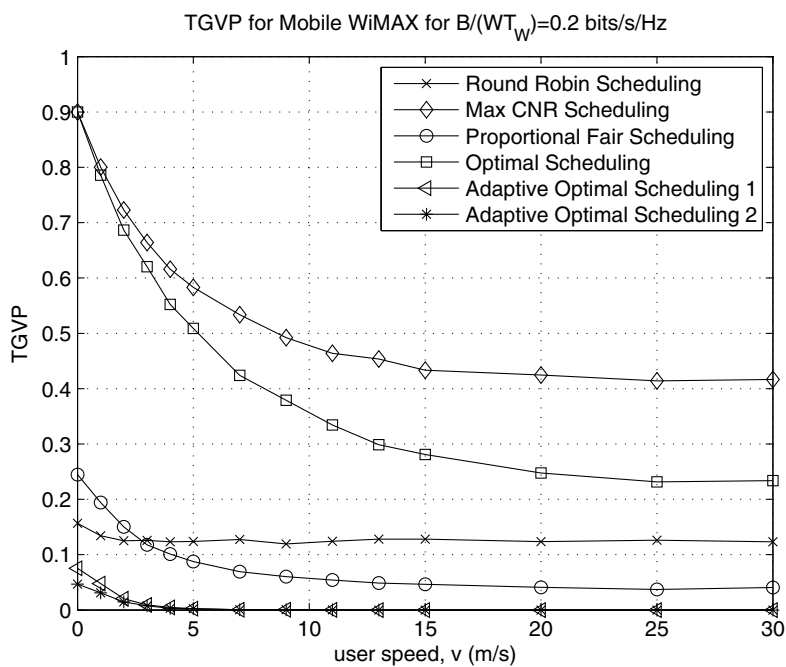

Figure 9 TGVP versus Speed for 10 users in a Mobile WiMAX network with identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz, where $T_W = 80$ ms corresponding to 16 time-slots. Each value in the plot is an average over 1,000 Monte Carlo simulations.
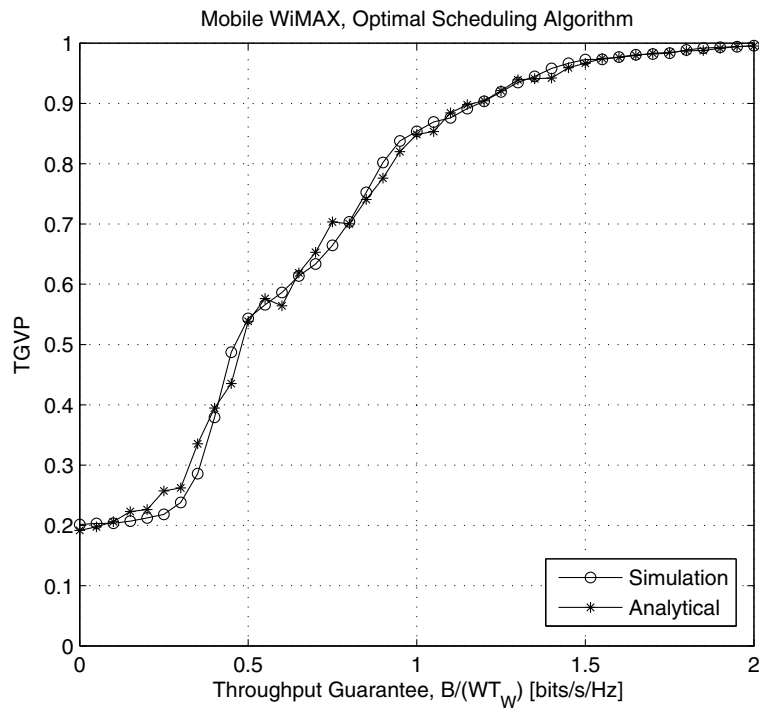
Figure 10 Approximated TGVP versus Monte Carlo simulated TGVP for a network with 10 users. Plotted for the Mobile WiMAX time-slot length of 5 ms and a time-window of $T_W = 80$ ms, corresponding to $K = 16$ time-slots. Each value in the simulated graph is an average over 1,000 Monte Carlo simulations.
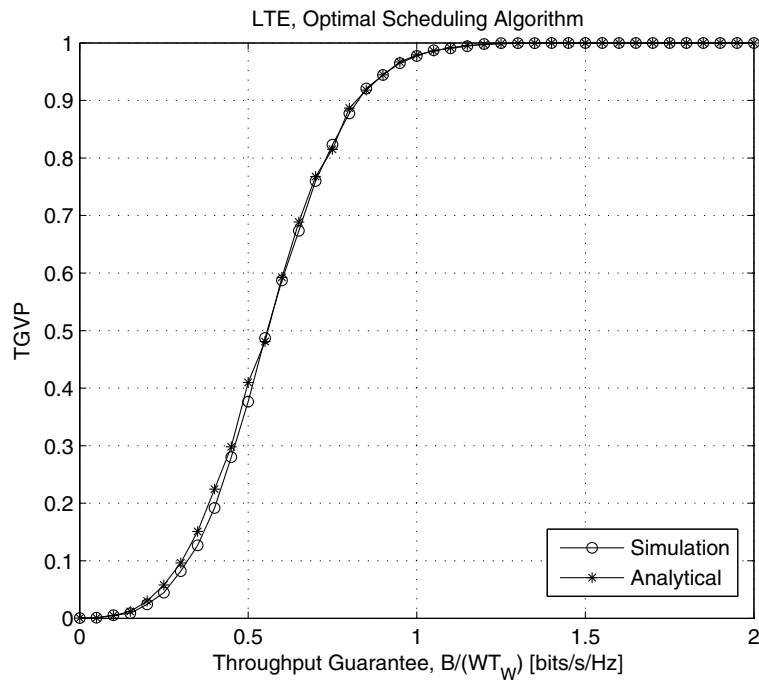

Figure 11 Approximated TGVP versus Monte Carlo simulated TGVP for a network with 10 users. Plotted for the LTE time-slot length of 1 ms and a time-window of $T_W = 80$ ms, corresponding to $K = 80$ time-slots. Each value in the simulated graph is an average over 500 Monte Carlo simulations.

are uncorrelated, while the Monte Carlo simulations are for users that have a correlated CNR from time-slot to time-slot. Jakes' correlation model is used with $f_c = 1$ GHz and a user speed of $v = 30$ m/s. Since the optimal algorithm is designed for long time-windows containing many time-slots, the TGVP for the Mobile WiMAX (with fewer time-slots) does not go to zero as B/WT_W approaches zero. However, as the number of time-slots increases (LTE), we observe that the TGVP of the optimal scheduling algorithm approaches zero.

The tightness of the approximation is both influenced by the number of time-slots K and the length of a time-slot T_{TS}. We need to calculate the TGVP for a relatively large number of time-slots K to obtain a tight approximation. For a fixed time-window T_W, larger K would mean shorter time-slots and we will thus experience a higher correlation between the time-slots. Since we have assumed uncorrelated time-slots to obtain our TGVP approximation, we will therefore have a less tight approximation for short time-slots. We see that the TGVP approximation for LTE network ($K = 80$ time-slots) is better than Mobile WiMAX ($K = 16$ time-slots). We can therefore conclude that the number of time-slots K within the time-window T_W will affect the tightness of the TGVP

approximation more than the fact that the shorter time-slots are more correlated.

Next, we observe the tightness of the TGVP approximation when we have longer time-window. Compare Figures 10 and 12 where the number of time-slots are 16 and 80, respectively. For high values of T_W, the value of K is higher and the correlation over the time-window is smaller. We can also conclude that long time-windows will lead to more tight TGVP approximations. Note also that the TGVP performance is better for the second case, since there are more time-slots now, and the optimal scheduling algorithm is designed for more time-slots.

10 Discussion

While much research has been done on providing *long-term* throughput guarantees, little study has addressed how to guarantee the *short-term* throughput to the users. The research in short-term performance has mainly focused on fairness issues. Our proposed adaptive scheduling algorithms are thus significant. In this section, we list some of the other studies that focus on short-term throughput guarantees.

In [26], the authors extend wireline scheduling policies to wireless networks and present wireless fair scheduling
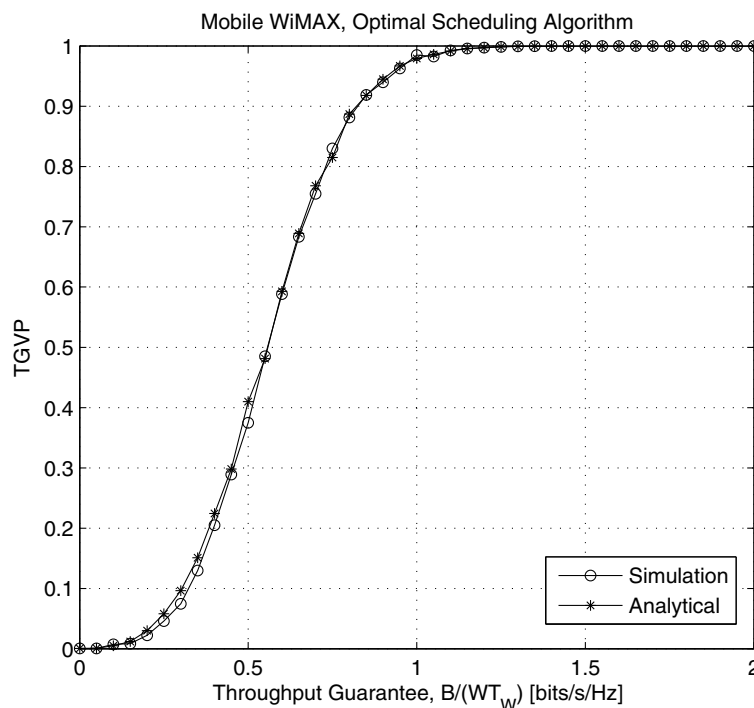

Figure 12 Approximated TGVP versus Monte Carlo simulated TGVP for a network with 10 users. Plotted for the Mobile WiMAX time-slot length of 5 ms and a time-window of $T_W = 400$ ms, corresponding to $K = 80$ time-slots. Each value in the simulated graph is an average over 1,000 Monte Carlo simulations.

policies which give short- and long-term throughput guarantee bounds.

The authors in [27] have analyzed and evaluated the problem of real-time users' short-term QoS probabilistic properties, in terms of maximum delay and minimum received throughput guaranties, under basic opportunistic scheduling policies (MCS and PFS). In [21], the authors argued that the probabilistic delay constraints are insufficient indicators of real-time QoS requirements, and probabilistic short-term throughput guarantees are more appropriate criteria. Based on this argument, they developed and evaluated a utility-based opportunistic resource allocation algorithm which aims at the minimization of real-time users' short-term TGVPs.

An algorithm aimed specifically at providing short-term throughput guarantees has been proposed in [28]. However, this algorithm achieves a significantly lower average long-term throughput as compared to MCS scheduler or PFS algorithm.

In [29], the authors proposed a predictive proportional fair algorithm and showed that its short-term throughput performance is better than the PFS algorithm.

11 Conclusions
For wireless networks carrying real-time traffic, providing throughput guarantees is interesting both from the customers' and the network providers' point of view. In order to have the most efficient utilization of the network, a scheduler in such a network should try to distribute the amount of bits that can be received or transmitted by each user according to given throughput guarantees. In this article, we have formulated an optimization problem which aims at finding the maximum number of bits that can be guaranteed to the users within a time-window for a given set of system parameters. By building on the results in [11] and by assuming that the distributions of the users' CNRs are known, we find an optimal scheduling algorithm, both for the case where the throughput guarantees are different from user to user, and for the case where the users have the same throughput guarantees. To further improve the short-term performance of this algorithm, we propose two adaptive and low complexity versions of the optimal algorithm. Results from our simulations show that the proposed adaptive algorithms perform significantly better than any of the other well-known scheduling algorithms in Mobile WiMAX-, HSDPA-, LTE-, and WINNER I-based networks. For systems that have many time-slots within the time-window, e.g., for WINNER I, the optimal scheduling algorithm also performs better than all the other well-known algorithms. For a network with heterogeneous throughput guarantees, the proposed adaptive scheduling algorithms are the only

algorithms that support a throughput guarantee close to unity, i.e., TGVP = 0. The second adaptive scheduling algorithm is simpler in implementation but still provides similar throughput guarantees as provided by the first algorithm. Furthermore, it is also independent of the time-window and therefore overcomes the problem of fixed time-window placement of the first algorithm. The simulations have also shown that the performance of the adaptive algorithms at lower users' speed (strong temporal correlation) also remains the best.

We have also derived an approximate expression for the TGVP which can be obtained in a time-slotted wireless network with any scheduling policy with a given set of system parameters, known cumulants of the bits transmitted to or from the scheduled user in a time-slot, and a given distribution of the number of time-slots allocated to a user within a time-window. From the simulations, we conclude that correlated time-slots have a small effect on the tightness of the approximations as compared to the number of time-slots. It can also be concluded that the TGVP approximations are tighter for relatively long time-windows.

The analysis in this study involves several idealistic assumptions. How realistic these assumptions are for real-life networks is a subject for further research.

Endnotes
[a]For an infinitely long time-window T_W, (15) and (16) become equalities. [b]Initially proposed by the authors in [13].

Author details
[1]Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim NO-7491, Norway [2]Telenor Corporate Development, Fornebu NO-1331, Norway [3]EMGS ASA, P.O. Box 2087 Vika, Oslo NO-0125, Norway

Competing interests
The authors declare that they have no competing interests.

References
1. F WiMAX, Mobile WiMAX–Part II: a comparative analysis. http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part2_Comparative_Analysis.pdf
2. R Knopp, PA Humblet, Information capacity and power control in single cell multiuser communications, in *Proceedings of IEEE International Conference on Communications (ICC'95)*, 331–335 (1995)
3. P Viswanath, DNC Tse, R Laroia, Opportunistic beam forming using dumb antennas. IEEE Trans Inf Theory. **48**(6), 1277–1294 (2002). doi:10.1109/TIT.2002.1003822

4. Y Liu, E Knightly, Opportunistic fair scheduling over multiple wireless channels, in *Proceedings of IEEE Joint Conference of the Computer and Communications Societies (INFOCOM'03)*, vol. 2 (San Francisco, CA,USA, 2003), pp. 1106–1115

5. V Hassel, MR Hanssen, GE Øien, Spectral efficiency and fairness for opportunistic round robin scheduling, in *Proceedings of IEEE International Conference on Communications (ICC'06)*, vol. 2 (Istanbul, Turkey, 2006), pp. 784–789

6. JM Holtzman, Asymptotic analysis of proportional fair algorithm, in *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'01)*, vol. 2 (San Diego, CA, USA, 2001), pp. F-33–F-37

7. D Avidor, S Mukherjee, J Ling, C Papadias, On some properties of the proportional fair scheduling policy, in *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, vol. 2. (Barcelona, Spain, 2004), pp. 853–858

8. X Liu, EKP Chong, NB Shroff, A Framework for Opportunistic Scheduling in wireless networks. Comput Netw. **41**(4), 451–474 (2003). doi:10.1016/S1389-1286(02)00401-2

9. M Andrews, L Qian, A Stolyar, Optimal utility based multi-user throughput allocation subject to throughput constraints, in *Proceedings of the 24th annual joint conference of the IEEE Conference Computer and Communications Societies (INFOCOM'05)*, vol. 4 (Miami, FL, USA, 2005), pp. 2415–2424

10. S Redana, A Frediani, A Capone, Quality of service scheduling based on utility prediction, in *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08), Cannes*, 1–5 (2008)

11. SC Borst, P Whiting, Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization. IEEE Trans Veh Technol. **52**(3), 569–586 (2003). doi:10.1109/TVT.2003.810967

12. V Hassel, S de la Kethulle de Ryhove, GE GE Øien, Scheduling algorithms for increased throughput guarantees in wireless networks, in *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS'07)*, Trondheim, Norway, 2007, pp. 401–406

13. J Rasool, GE Øien, A simplified adaptive scheduling algorithm for increased throughput guarantees, in *Proceedings of 2010 European Wireless Conference (EW2010)*, Lucca, Italy, 2010, pp. 182–189

14. J Rasool, GE Øien, Quantifying the throughput guarantees offered in wireless networks, in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'11)*, Cancun, Mexico, 2011, pp. 1173–1178

15. R Suoranta, KP Estola, S Rantala, H Vaataja, PDF estimation using order statistic filter bank, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'94)*, vol. 3. Adelaide, Australia, 1994, pp. III-625–III-628

16. V Hassel, GE Øien, D Gesbert, Throughput guarantees for wireless networks with opportunistic scheduling: a comparative study. IEEE Trans Wirel Commun. **6**(12), 4215–4220 (2007)

17. JA Bucklew, Introduction to Rare Event Simulation, Springer Series in Statistics (Springer, New York, 2004)

18. Hugo Touchette, Legendre-Fenchel transforms in a nutshell, http://www.maths.qmul.ac.uk/~ht/archive/lfth2.pdf (August 2007)

19. L Yang, MS Alouini, Performance analysis of multiuser selection diversity, in *IEEE International Conference on Communications (ICC'04)*, Paris, France, 2004, pp. 3066–3070

20. E Kreyszig, *Advanced Engineering Mathematics*, 7th edn. (Wiley, New York, 1993)

21. T Kastrinogiannis, S Papavassiliou, Utility based short-term throughput driven scheduling approach for efficient resource allocation in CDMA wireless networks. Wirel Personal Commun J. **52**(3), 517–535 (2010). doi:10.1007/s11277-008-9632-9

22. F Fluckiger, *Understanding Networked Multimedia, Applications and Technology*, (Prentice-Hall, Inc., Englewood Cliffs, 1995)

23. M Sternad, T Svenson, G Klang, WINNER MAC for cellular transmission, in *Proceedings of IST Mobile Summit*, Mykonos, Greece, 2006

24. Freescale Semiconductor, Long Term Evolution Protocol Overview, http://www.freescale.com/files/wireless_comm/doc/white_paper/LTEPTCLOVWWP.pdf (2008)

25. HJ Bang, Advanced scheduling techniques for wireless data networks, Master Thesis (Department of Physics, University of Oslo, Norway, 2005)

26. S Lu, V Bharghavan, R Srikant, Fair scheduling in wireless packet networks. IEEE/ACM Trans Netw. **7**(4), 473–489 (1999). doi:10.1109/90.793003

27. T Kastrinogiannis, S Papavassiliou, Probabilistic short-term delay and throughput requirements of multimedia services in high throughput wireless networks, in *Proceedings of IEEE Sarnoff Symposium on Advances in Wired and Wireless Communications* (Nassau Inn, Princeton, NJ, 2007), pp. 1–5

28. N Chen, S Jordan, Downlink scheduling with guarantees on the probability of short-term throughput. IEEE Trans Wirel Commun. **8**(2), 593–598 (2009)

29. HJ Bang, T Ekman, D Gesbert, Channel predictive proportional fair scheduling. IEEE Trans Wirel Commun. **7**(2), 482–487 (2008)