Resource Allocation Under Uncertainty Using the Maximum Entropy Principle

Mathias Johansson, *Member, IEEE*, and Mikael Sternad *Senior Member, IEEE*

Abstract— **In this paper we formulate and solve a problem of resource allocation over a given time horizon with uncertain demands and uncertain capacities of the available resources.**

In particular, we consider a number of data sources with uncertain bit rates, sharing a set of parallel channels with time varying and possibly uncertain transmission capacities. We present a method for allocating the channels so as to maximize the expected system throughput. The framework encompasses quality-of-service requirements, e.g. minimum-rate constraints, as well as priorities represented by a user-specific cost per transmitted bit.

We assume only limited statistical knowledge of the source rates and channel capacities. Optimal solutions are found by using the maximum entropy principle and elementary probability theory.

The suggested framework explains how to utilize multiuser diversity in various settings, a field of recently growing interest in communication theory. It admits scheduling over multiple base stations and includes transmission buffers to obtain a method for optimal resource allocation in rather general multiuser communication systems.

Keywords— **Maximum Entropy, Multiuser Diversity, Resource Allocation, Scheduling, Uncertainty.**

I. Introduction

In this paper we consider a problem of allocating bandwidth among users sharing a number of channels. A number of sources are producing bits at unknown rates. These bits are to be transmitted to U users (or receivers). The sources share a number, R, of transmission channels (or resources) which may be used to send the produced bits to the receivers.

The problem is an extension and reformulation of a simpler resource allocation problem, the "widget problem", studied by Jaynes [1] (also mentioned in [2] ch. 14), where there were three possible decisions and one resource, with known constant capacity, which could only be used exclusively for one task.

In our present problem each channel-receiver pair has a time-varying number associated with it, denoting the number of bits that can be sent over the link at a prescribed bit error rate (BER), given that the channel is used exclusively for transmitting to that specific receiver. We will henceforth denote this number as the *effective capacity*[1] of that link.

Bits produced by the sources are stored in buffers monitored by a transmission controller. The transmission controller aims to distribute the bits over the channels so that the number of bits in the buffers is minimized, or equivalently so that the system throughput is maximized. The question that we address is then: given only limited knowledge of the actual source rates and effective capacities, how should the controller distribute the resources?

The main information-theoretic motivation for using scheduling in mobile communications comes from the observation [3] that the sum-of-rates capacity for a single channel increases with the number of users and that it is maximized by transmitting exclusively to the user with highest channel power. This phenomenon, denoted multiuser diversity [4], suggests that independent channel fluctuations between different users should be taken advantage of instead of being combatted. The concept is very similar to multi-antenna diversity. Knopp [4] describes it as selection diversity at the transmitting end. The result in [3] however assumes perfect channel knowledge, a single channel, additive Gaussian disturbances only, and that transmission buffers cannot be emptied.

Following the publication of [3], scheduling in wireless communications has received an increasing amount of attention, but the focus has been on assuming that there is always data to send (buffers are never emptied) and that the scheduler has perfect channel knowledge.

In high-level schedulers, stochastic channels are sometimes introduced by two-state models (error-free or random errors) [5], which might be considered too coarse. In [6], [7] a framework is suggested for scheduling several timeslots ahead which takes known buffer sizes into account but requires perfect channel prediction. Another rule, the proportional fair scheduler [8], gives exclusive access to the user who currently has the highest effective capacity normalized by its average allocated throughput, thus striking a balance between fairness and performance, but again requiring complete knowledge of the effective capacities. A similar result to that in [3] is obtained in [9] for a set of parallel broadcast channels corrupted only by additive white Gaussian noise. Another line of work [10], [11] which has been used for multi-hop networks and on-off types of links with constant effective capacity considers queue stability as the main criterion. An interesting application of this criterion which also shows a relation to the proportional fair scheduler is reported in [12], where queue stabilizing schedulers are adapted to support quality-of-service (QoS) constraints.

Until recently, little had been published concerning al-

Mathias Johansson is with Dirac Research AB and the Signals & Systems Group, Uppsala University, Uppsala, Sweden. E-mail: mathias.johansson@signal.uu.se .

Mikael Sternad is with the Signals & Systems Group, Uppsala University, Uppsala, Sweden. E-mail: mikael.sternad@signal.uu.se .

[1]The term capacity is here used in a non-traditional way and should not be confused with any of the usual information theoretic capacity definitions. The effective capacity denotes the transmission rate for a given BER requirement that a user obtains if no other users transmit simultaneously on the channel. The actual transmission rate becomes less than that if the channel is shared among several users.

location of multiple shared transmitters except for base station assignments in the uplink with the objective of minimizing allocated mobile powers [13], [14] and a similar downlink problem [15]. After the submission of this paper however, the capacity region for both the Gaussian multiple-input multiple-output (MIMO) broadcast and multiple-access channel has been found under the assumption of perfect channel knowledge at the transmitter and subject to the qualification that there is always data to send [16], [17], [18]. In order to achieve the sum capacity for any of these MIMO channels, exclusive allocations must in general be abandoned. Moreover, all currently known schemes require substantial amounts of channel feedback and are extremely computationally demanding. In [19] the case of partial channel knowledge is investigated and a simplified resource allocation scheme based on using several randomized beams is devised. The scheme transmits to the user with maximum signal-to-interference ratio on each beam. When the number of users approaches infinity, this scheme approaches the capacity-optimal scheme. For the case of few users, however, there is still a lack of low-complexity low-feedback schemes that approach the sum capacity.

In this paper, we do not focus on the general MIMO case; the model considered here assumes parallel channels where use of one channel does not affect any other channel. Instead, our focus is on scheduling transmissions under uncertain channel conditions and uncertain source rates with the objective of maximizing total throughput under quality-of-service constraints. These topics have hitherto not been investigated in any detail. The aim of this paper is to provide such a study.

In summary, this work extends the current literature by providing means for resource allocation with uncertain source rates, taking buffer levels into account, and scheduling with multiple parallel transmitters over arbitrary time periods. Furthermore, the scheduling framework is extended to take into account inaccurate channel predictions.

In two seminal papers [20], [21] , Jaynes introduced the maximum entropy principle as a consistent method for determining probability distributions under constraints on mean values of functions of data. The principle is applicable to inference problems with well-defined hypothesis spaces but incomplete data. A motivation for its use is contained in the *entropy concentration theorem* [22], which states that given the imposed constraints, the maximum entropy distribution can be realized in overwhelmingly more ways than any other distribution. It is thus considered as the least biased solution for determining prior probabilities under the given constraints. It has been successfully applied to a variety of problems, the reference list providing a sample of examples from image reconstruction [23] [24], spectrum estimation [25], finance [26], language modelling [27], and physics [28], [29]. We here propose that the maximum entropy principle be used for modelling uncertain data flows in mobile communications systems.

The paper is organized as follows: in Section II we present the problem formulation, whereas in Section III

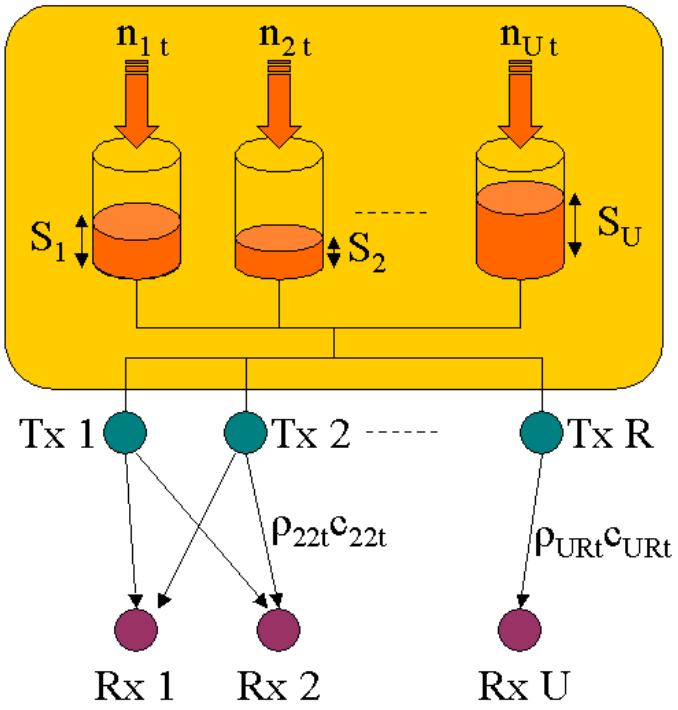


Fig. 1. The system consists of U buffers, one for each receiver. R transmission resources are available and user u receives $\rho_{urt}c_{urt}$ bits at time t from transmitter r.

we recapitulate the maximum-entropy principle and use it to model the uncertain source flows. Following this, Section IV presents the solutions for three different states of knowledge concerning source rates and effective capacities. Before concluding the paper, in Section V some observations are made concerning the behavior of the scheduler for different degrees of uncertainty. The performance is also compared to that obtained by the proportional fair scheduler.

II. Distributing Bandwidth Among Users Sharing a Set of Channels

The problem we shall investigate is how to allocate transmission resources with possibly uncertain effective capacities to sources with uncertain bit rates. A motivating application has been the problem of link-level predictive scheduling of a broadband downlink radio resource to mobile users with independently varying channel capacities due to fast fading [30], [6]. Here we consider a slightly generalized problem.

In Figure 1 an overview of the system is given. There are U users, and equally many buffers. We will schedule the use of the R channels for T time slots. Each channel is here taken to be a transmission resource that is orthogonal to all other channels, ie. usage of one channel does not affect other effective channel capacities. For instance, a channel may be a frequency bin in an OFDMA system or one out of several non-interfering antenna beams.

During the scheduling horizon T, each buffer is filled with n_u bits, u denoting the user index. A buffer may also

have a number, S_u, of bits remaining in stock from previous scheduling rounds. The objective of interest will be to minimize the buffer contents at the end of the scheduled time horizon. In a completely deterministic situation, this amounts to minimizing the loss function

$$L = \sum_{u=1}^{U} g\big(S_u + n_u - \sum_{t=1}^{T}\sum_{r=1}^{R} c_{urt}\rho_{urt}\big) \quad , \tag{1}$$

where $g(x) = x$ if $x > 0$, else $g(x) = 0$. The time-varying effective capacity of the rth channel to user u is denoted by the integer c_{urt}, while ρ_{urt} is the fraction ($0 \le \rho_{urt} \le 1$) of the bandwidth of the rth channel that we allocate to user u at time t. For instance, if $\rho_{urt} = 1$, user u uses the rth channel exclusively at time t. The total channel usage $\sum_u \rho_{urt}$ for a given channel r at a time t must satisfy $\sum_u \rho_{urt} \le 1$. The minimization of (1) would be performed by adjusting ρ_{urt} under whatever constraints the specific system poses on ρ_{urt}.

The total number of incoming bits, n_u, in the time interval T is the sum of the influxes at each time instant t:

$$n_u = \sum_{t=1}^{T} n_{ut} \quad . \tag{2}$$

In cases where we have knowledge of time variations, we will use this more detailed notation. In general, as a notational convention, for any quantity a, we will use at most three indices: a_{urt}, where u ($1 \le u \le U$) denotes user index, r ($1 \le r \le R$) channel index, and t ($1 \le t \le T$) time index. Whenever any of these three indices is omitted the quantity represents the sum over all values of the omitted index.

In general, complete knowledge of the effective capacities or the number of incoming bits at any specific future time is unavailable. Therefore we cannot directly minimize L but must resort to assigning probability distributions for n_u and c_{urt} and minimize the expected loss. Assuming that knowledge of effective channel capacities gives no information of incoming bit rates [2], and vice versa, we can factor the joint prior probability[3]

$$\begin{aligned} P(n_u c_{urt}|I) &= P(n_u|c_{urt},I)P(c_{urt}|I) = \\ &= P(n_u|I)P(c_{urt}|I) \end{aligned} \tag{3}$$

and the expected loss becomes

$$\begin{aligned} \langle L\rangle &= \sum_{u=1}^{U}\sum_{c_{urt}=0}^{\infty}\sum_{n_u=0}^{\infty} P(n_u|I)P(c_{urt}|I) \times \\ &\times \; g\left(S_u + n_u - \sum_{t=1}^{T}\sum_{r=1}^{R} c_{urt}\rho_{urt}\right) \quad . \end{aligned} \tag{4}$$

[2]Although certain communication protocols actually change their transmission rates due to channel variations, these protocols, eg. TCP (Transmission Control Protocol), react on slower time scales than would normally be used in scheduling decisions at the link layer.

[3]To indicate that the probability expressions will change according to the information at hand, all probabilities are conditioned on I, which denotes any available information relevant for inferring n_u or c_{urt}.

Throughout the rest of the paper we will find it convenient to use the notation $\langle L_u\rangle$ for the expected loss contribution corresponding to user u, with the total expected loss being the sum of all user contributions:

$$\langle L\rangle = \sum_{u=1}^{U}\langle L_u\rangle \quad . \tag{5}$$

The scheduling framework we propose relies on minimizing (4) subject to various constraints. The rest of the paper is concerned with deriving the expected loss contributions $\langle L_u\rangle$ for a few typical cases in mobile communications. It should be emphasized that the cases differ only in what knowledge the scheduler uses.

Finding the minimum of (4) will in general turn out to require non-linear programming. The basic constraints on ρ_{urt} are:

$$\sum_u \rho_{urt} \le 1 \quad \forall r,t \tag{6}$$

$$0 \le \rho_{urt} \le 1 \quad \forall u,r,t \; , \tag{7}$$

but in general we may have an additional number of matrix equalities and inequalities representing constraints imposed by the specific system architecture on different resources. Examples of such constraints include

- a limited set Ω of rate levels, implying that the transmission rate $\rho_{urt}c_{urt}$ must belong to the set Ω,
- in a time division system, ρ_{urt} can only be 0 or 1,
- some channels may not be accessible to all users, i.e. $\exists r, \exists u,\ \rho_{urt} = 0$,
- in a network guaranteeing some minimum level of service quality, constraints may take the form of user-specific minimum channel access levels, $\rho_{urt} \ge \eta_{urt}$, or minimum transmission rates $\sum_r \sum_t \rho_{urt}c_{urt} \ge \varphi_u$.

These types of constraints are readily treated by available software for solving non-linear programming problems and present no conceptual difficulties. The general problem can thus be transformed to different specialized settings, all represented by the same average loss function but with different optima due to the restrictions on ρ_{urt}.

Minimizing the number of bits remaining in stock is equivalent to maximizing the sum of the users' bit rates. With this criterion, user specific priorities can be introduced as multipliers to each user's loss contribution in (5). This can be interpreted as a user-specific cost per bit, expressed as a function $\pi(u,\{\theta_u\})$ of any set $\{\theta_u\}$ of known parameters (such as time, delay, buffer levels, average effective capacities, average influxes, bit prices, etc.). The generalized criterion is then to minimize

$$\langle L\rangle = \sum_{u=1}^{U}\pi(u,\{\theta_u\})\langle L_u\rangle \quad . \tag{8}$$

For instance, if $\pi(u,\{\theta_u\})$ is defined as the reciprocal of user u's average throughput, we obtain a generalized version of the proportional fair scheduler [8]. We will not consider fairness any further; it is sufficient to note that any fairness

requirement or user priority that can be formulated as a deterministic function describing an equivalent user-specific cost per bit is compatible with the given formulation.

Another possible approach could be to use quadratic criteria in order to punish large buffers and consequently aim at reducing the risk of buffer overflow. A disadvantage of using a quadratic criterion here is that the scheduler would no longer maximize the sum of the users' bit rates, hence capacity would be wasted. Another problem is that if priorities are introduced as multiplicative factors for each user's contribution to the total loss, the priorities will loose their intuitive meaning as incurring a certain cost per bit to the network. It can be shown that some queue stabilizing schedulers are local approximations to using a quadratic criterion on the buffer levels (see [31] and [32]). Thus, they do not maximize throughput and have a risk of starving other users when a single user floods its buffer.

In the sections following the next we derive the expected loss contribution for each user u, $\langle L_u \rangle$ for different states of prior information by the use of the maximum entropy principle. Solutions are given for three different states:

- Section IV-A assumes knowledge of *average* source rates and *exactly* known capacities.
- In Section IV-B we relax the requirement of perfect channel knowledge and instead assume *capacity predictions of varying accuracy.*
- Finally, in Section IV-C source flows are subdivided into packets and the scheduler requires knowledge of the *average number of packets produced for each packet size.*

III. The Maximum Entropy Approach to Source Flow Modelling

Building on Shannon's explanation [33] of entropy for discrete events,

$$H = -\sum_A P(A|I) \log P(A|I) \quad , \tag{9}$$

as a measure of uncertainty[4], Jaynes proposed [20], [21] that prior probabilities be constructed by maximizing the entropy under the constraints given by the information at hand. The solution is considered to be the least biased possible as any other solution would imply lower entropy and thus lead to a less uncertain state than implied by the given information. In effect, unwarranted assumptions, or information that is not available, would be injected into the consequent inference. The *entropy concentration theorem* [22] further establishes that the maximum entropy distribution is the sampling distribution which can arise in the greatest number of ways under the imposed constraints. Specifically, if in a long data sequence certain mean values of the sequence have been recorded but not the actual sequence itself, then out of all possible sequences that satisfy the given mean values, the overwhelming majority will have an entropy extremely close to the maximum. This is a combinatorial fact similar to the asymptotic equipartition principle [33]; the longer the sequence and the more mean values recorded, the tinier the fraction of sequences that does not follow the maximum entropy distribution. A main motivation for using maximum entropy distributions is thus simply that there are so many more of them! This has been taken to mean that use of the maximum entropy principle for assigning priors under incomplete information results in a "discipline for avoiding unnecessary assumptions" [34]. Formal properties of maximum entropy distributions are given in [2], ch. 11.

The source flows in the current problem are not assumed to be known in detail. A common assumption concerning near-future networks is that traffic to a large extent will consist of Internet flows. Modelling an individual Internet data source is however a notoriously difficult problem [35]. Various distributions have been proposed, the most commonly used consists of assuming that the number of packets per time unit is Poisson distributed. This distribution has some justification when the incoming packet streams stem from a large number of independent sources, but not in the case of a single-user source flow. Another approach would be to record individual histograms for each user in the transmitter and use them as approximate probability distributions. That is however not realistic; the amount of data that has to be collected would typically be larger than that obtainable during a user's connection.

Instead, we propose to use the maximum entropy approach. We shall use the maximum entropy principle to model the source rates n_u subject to knowledge of the average source rate $\langle n_u \rangle$ for each user[5]. We first recapitulate the general maximum entropy problem and its solution, and then derive the distribution for the source rates.

A. Finding a maximum entropy distribution

Consider a problem where we have knowledge of mean values F_k of certain functions, $f_k(\cdot)$, of data:

$$\sum_{i=1}^{n} P_i f_k(x_i) = F_k \quad , \qquad 1 \le k \le m \tag{10}$$

where P_i denotes the probability for each possible "state of nature", indexed by $i \in \{1..n\}$.

We wish to find the set of probabilities P_i, for all possible i, that maximizes the entropy

$$H = -\sum_{i=1}^{n} P_i \log P_i \quad . \tag{11}$$

This is a standard variational problem solvable by using Lagrange multipliers when $m < n$. In Appendix A it is shown that using the partition function [20]

$$Z(\lambda_1, \ldots, \lambda_m) \equiv \sum_{i=1}^{n} \exp\left[-\lambda_1 f_1(x_i) - \ldots - \lambda_m f_m(x_i)\right] \tag{12}$$

[4]Although the logarithm in the entropy expression may be taken to any base, in this paper we restrict log to denote the natural logarithm.

[5]The average source rate can be estimated at the transmitter based on the incoming data. An initial estimate can be obtained by using the average of all users' data streams.

we have the formal solution

$$P_i = \frac{1}{Z(\lambda_1, \ldots, \lambda_m)} \exp\left[-\lambda_1 f_1(x_i) - \ldots - \lambda_m f_m(x_i)\right] \ , \tag{13}$$

where $\{\lambda_k\}$ are the Lagrange multipliers which are chosen so as to satisfy the constraints (10). This is the case when

$$F_k = -\frac{\partial}{\partial \lambda_k} \log Z, \quad 1 \leq k \leq m \ . \tag{14}$$

In (10) - (14) we have the general maximum entropy problem and its solution. It should be noted that the solution presented here automatically includes the constraint $\sum_{i=1}^{n} P_i = 1$ without need for an additional Lagrange multiplier.

B. The maximum-entropy probability distribution for the source flows

We are to assign a prior probability distribution for non-negative integer quantities, n_u, $u = 1 \ldots U$, having known means $\langle n_u \rangle$. Denoting this information by I, we now turn to find the $P(n_u|I)$ which maximizes the entropy $-\sum_{n_u} P(n_u|I) \log P(n_u|I)$ under the constraints

$$\langle n_u \rangle \ = \ \sum_{n_u=0}^{\infty} n_u P(n_u|I) \ , \ u = 1 \ldots U \ . \tag{15}$$

Notice that the summation index reflects that the integer n_u is non-negative. The partition function (12) becomes

$$\begin{aligned} Z(\lambda_1, \ldots, \lambda_U) &= \sum_{n_1=0}^{\infty} \ldots \sum_{n_U=0}^{\infty} \exp(-\lambda_1 n_1 - \ldots - \lambda_U n_U) \\ &= \sum_{n_1=0}^{\infty} \left(\ldots \left(\sum_{n_U=0}^{\infty} \exp(-\lambda_U n_U) \right) \ldots \right) \exp(-\lambda_1 n_1) \\ &= \prod_{u=1}^{U} \frac{1}{1 - e^{-\lambda_u}} \ , \end{aligned} \tag{16}$$

where we first rewrote the expression according to $x^{a+b} = x^a x^b$ and then used the closed form expression for the geometric series. The Lagrange multipliers are now determined from (14):

$$\langle n_u \rangle = -\frac{\partial}{\partial \lambda_u} \log Z = \frac{1}{e^{\lambda_u} - 1} \ . \tag{17}$$

Independence between different probabilities yields higher entropy than dependencies, and consequently the maximum-entropy probability assignments $P(n_u|I)$ factor:

$$P(n_1, \ldots, n_U|I) = P(n_1|I) \ldots P(n_U|I) \ . \tag{18}$$

Inserting (16) into (13) and using (18) and (17) we obtain

$$\begin{aligned} P(n_u|I) &= (1 - e^{-\lambda_u}) e^{-\lambda_u n_u} \ , \ n_u = 0 \ldots \infty \\ &= \frac{1}{\langle n_u \rangle + 1} \left(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1} \right)^{n_u} \end{aligned} \tag{19}$$

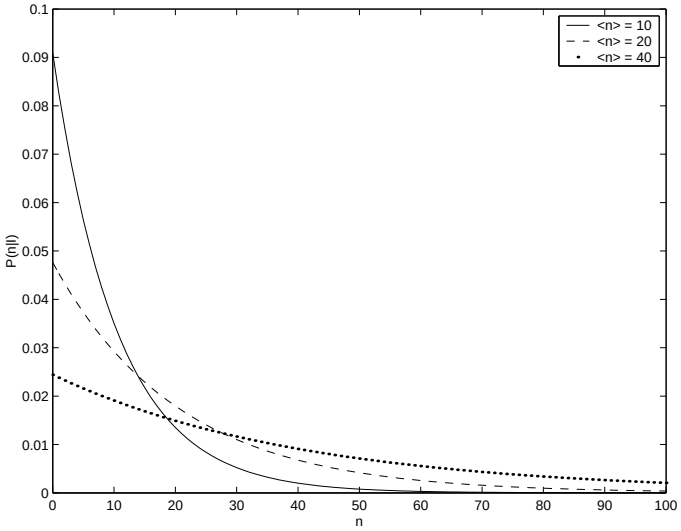


Fig. 2. The maximum entropy probability distribution for a non-negative integer quantity n with known mean $\langle n \rangle$.

as the distribution of highest entropy subject to the constraints (15) and $\sum P(n_u|I) = 1$.

The maximum-entropy derivation of the negative exponential distribution above can also be found in [1]. In Figure 2 the distribution is plotted for different mean values. The skewness of the curve arises because n_u is only defined for non-negative values. Hence, for a larger mean value the curve tends more and more towards a uniform distribution. The distribution would be different if n_u had a known upper bound. For instance, if the n_u represent the number of dots on the face of a die, we must include that $1 \leq n_u \leq 6$ in our probability derivation. This yields a distribution which is skewed differently depending on the given mean values[6].

IV. Specific Solutions to the General Resource Allocation Problem

A. Knowledge of average source rates and exact capacities

Here we will work out the expected loss contribution of user u, $\langle L_u \rangle$ (cf. (5)), for the scheduling problem when the average number of incoming bits during the interval T, $\langle n_u \rangle$, in each buffer is known and the effective capacities of the transmitters are exactly known. Denoting this information by I and following the derivation in Section III-B we assign $P(n_u|I)$:

$$P(n_u|I) = \frac{1}{\langle n_u \rangle + 1} \left(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1} \right)^{n_u} \ . \tag{20}$$

For clarity, we introduce

$$x_u = \sum_{t=1}^{T} \sum_{r=1}^{R} c_{urt} \rho_{urt} \ , \tag{21}$$

[6]For the case of data flows, there is an upper bound which is determined by the bandwidth of the fixed network preceding the buffers. This limit is neglected here because it is usually much larger than the expected source flow of each user.

describing the total number of bits sent from buffer u over the scheduled time horizon T. With $P(n_u|I)$ given by (20) the expected loss contribution with known c_{urt} becomes:

$$\langle L_u \rangle = \sum_{n_u=0}^{\infty} P(n_u|I) g(S_u + n_u - x_u) \quad (22)$$

$$= \begin{cases} \langle n_u \rangle \left(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1} \right)^{x_u - S_u} & , \; x_u > S_u \\ S_u + \langle n_u \rangle - x_u & , \; x_u \leq S_u \end{cases} . \quad (23)$$

The summation over n_u in (22) is evaluated in Appendix B with the result (23).

In certain problems the expected values of the influxes at time t defined in (2), n_{ut}, vary over time, i.e. we have knowledge of $\langle n_{ut} \rangle$ (defined analogously to (15)) for specified times t. In Appendix C the solution for this case is derived. The resulting loss contribution for time-varying expectations of incoming bit rates is:

$$\langle L_u \rangle = \begin{cases} K_u \langle n_{u1} \rangle \left(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1} \right)^{x_u - S_u} & , \; x_u > S_u \\ S_u + \langle n_u \rangle - x_u & , \; x_u \leq S_u \end{cases} , \quad (24)$$

with

$$K_u = \prod_{k=2}^{T} \frac{1}{\langle n_{uk} \rangle + 1} \times \frac{1}{1 - \frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1} \frac{\langle n_{u1} \rangle + 1}{\langle n_{u1} \rangle}} \quad (25)$$

where, for the case $x_u > S_u$, the influx averages in K_u are no longer ordered chronologically, but have been reordered by decreasing size, with the index k, to ensure convergence of the geometrical series. Notice also that K_u, the product over all averages which are smaller than $\langle n_{u1} \rangle$, is a constant that does not depend on the actual resource allocation ρ_{urt}. Therefore, if the minimum loss is calculated iteratively this factor need not be recalculated at each iteration.

B. Knowledge of average source rates and accuracy of capacity predictions

In this section we turn to a case which is of particular interest in applications for mobile communications. Here, a transmitter may predict future channel conditions with some known accuracy based on measured fading patterns at the receivers (see e.g. [36], or [37]). Adaptive modulation is then used to adjust the effective capacity.

We must now consider three different effective capacities: the *predicted* one $\hat{c}_{urt}$, the *potential* one $\bar{c}_{urt}$, and the eventually *obtained* one c_{urt}. The potential effective capacity $\bar{c}_{urt}$ is the number of bits that could be sent over the channel at time t with a prescribed error rate if we knew the channel and thus could choose the optimal modulation level. With inaccurate channel knowledge however, if the potential effective capacity is lower than predicted, then the modulation level may be set too high leading to a performance degradation due to increasing bit error rates. If on the other hand the predicted capacity is lower than the potential capacity, then the modulation level is set lower than the optimum and the *obtained* effective capacity will equal the predicted capacity (i.e. the obtained capacity will again be lower than the *potential* capacity). Thus, the probability for the outcome of the prediction (in the sense of being larger than, smaller than, or equal to the potential capacity) will determine the probability for obtaining a given effective capacity.

We assume that the accuracy of prediction is represented by a known variance, σ^2_{urt}, and that the prediction itself $\hat{c}_{urt}$ is the expected value of the potential (but unknown) effective capacity, $\bar{c}_{urt}$. As an example of how the prediction can be obtained, in [36], [37] an unbiased quadratic channel power predictor is derived, based on which it is possible to derive a pdf for the channel power ([36] ch. 7-8). Using that pdf one can determine the corresponding pdf for the effective capacity given a certain BER requirement by a change of variables. This can for instance be carried out by using the approximate BER expressions from [38]. Consequently, the expectation of the resulting pdf provides an unbiased prediction of the effective capacity.

In the case of a nonnegative integer quantity such as the potential effective capacity, finding the maximum-entropy distribution for known expectation and variance is analytically intractable. However, it is well-known [33] that the Gaussian distribution has the highest entropy for a given mean and variance if the quantity of interest is defined over the entire real axis. If the expectation of a Gaussian distribution is positive and large compared to its standard deviation, then it has negligible probability mass for negative numbers. Therefore, for reasonably accurate predictions of $\bar{c}_{urt}$ we may safely assign a Gaussian distribution as an accurate description of our state of knowledge.

However, as mentioned, the obtained capacity depends on the prediction error $\hat{c}_{urt} - \bar{c}_{urt}$. There are three possible cases:

1. $\underline{\hat{c}_{urt} \leq \bar{c}_{urt}}$. In this case the obtained effective capacity will equal the predicted one, $c_{urt} = \hat{c}_{urt}$.
2. $\underline{\bar{c}_{urt} \leq \hat{c}_{urt} \leq c^*_{urt}}$. If the predicted value is higher than the potential effective capacity, then the modulation level will be set too high and thus the obtained effective capacity will decrease. Here, c_{urt} is given by a function $f(\hat{c}_{urt})$ which depends on coding and other system-specific parameters. A reasonable approximation is to assume that the obtained effective capacity decreases linearly with the predicted value, reaching zero at a point $c^*_{urt} = v\bar{c}_{urt}$. We comment further on this model choice and the determination of v in the end of this section.
3. $\underline{\hat{c}_{urt} \geq c^*_{urt}}$. In this interval, the obtained capacity is zero.

In summary we obtain an effective capacity curve as described by Figure 3.

In Appendix D the probability for the obtained effective capacity c_{urt} given the predicted value is derived as the sum of the contributions from each of the three cases. It is shown that the probability for the obtained capacity is

$$P(c_{urt}|I) = P_1(c_{urt}|I) + P_2(c_{urt}|I) + P_3(c_{urt}|I) \quad (26)$$

where

$$P_1(c_{urt}|I) = \frac{1}{2}\delta(c_{urt} - \hat{c}_{urt}) \quad (27)$$

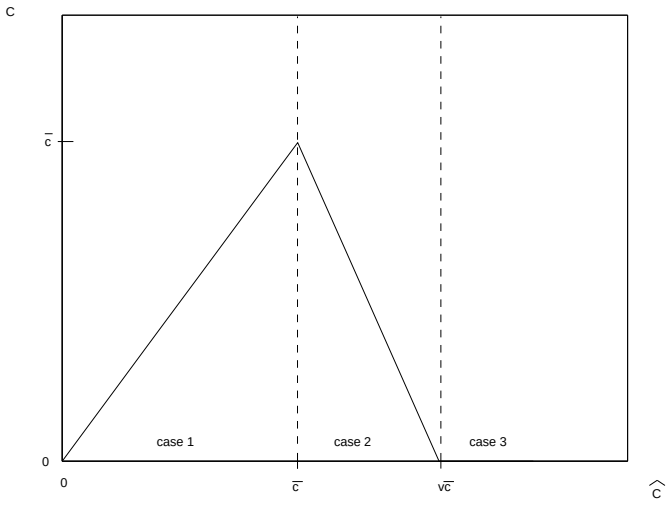


Fig. 3. The obtained capacity as a function of the predicted capacity with linear decline for too large predictions.

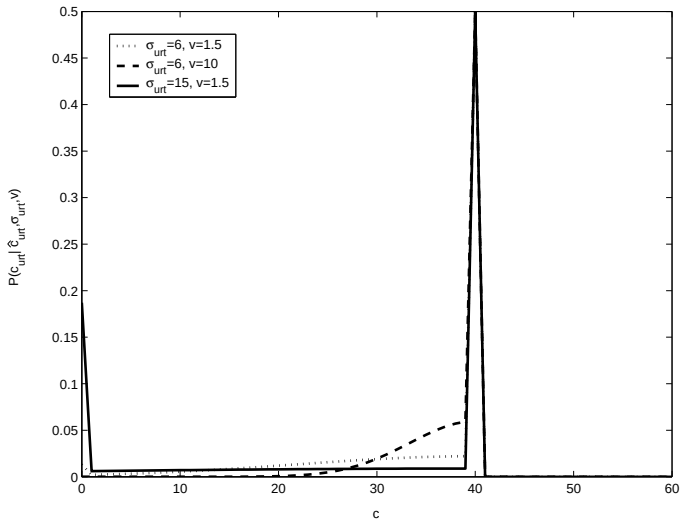


Fig. 4. The probability distribution for the obtained capacity given the prediction $\hat{c}_{urt} = 40$.

$$P_2(c_{urt}|I) = \frac{v-1}{\sqrt{2\pi}\sigma_{urt}v}\exp\left[-\left(\frac{v-1}{\sqrt{2}\sigma_{urt}v}\right)^2(c_{urt}-\hat{c}_{urt})^2\right] \times (H(c_{urt}) - H(c_{urt}-\hat{c}_{urt})) \quad (28)$$

$$P_3(c_{urt}|I) = \delta(c_{urt})\left(\frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{(v-1)\hat{c}_{urt}}{v\sigma_{urt}\sqrt{2}}\right)\right) \quad (29)$$

where $H(x)$ denotes the Heaviside step function and $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. The probability distribution for the obtained capacity is plotted for $\hat{c}_{urt} = 40$ and for different values of σ_{urt} and v in Figure 4.

We will now calculate each user's contribution $\langle L_u\rangle$ to the expected loss (4) with respect to $P(n_u|I)$ and $P(c_{urt}|I)$. Assuming independence between the two probability distributions, we can use the results obtained in the last section. The expected loss contribution will consist of a sum of two components, one for $x_u > S_u$ and another for $x_u \leq S_u$, weighted by their respective probabilities $P(x_u > S_u|I)$ and $1 - P(x_u > S_u|I)$:

$$\langle L_u\rangle = P(x_u > S_u|I)\langle L_{u1}\rangle + (1 - P(x_u > S_u|I))\langle L_{u2}\rangle \quad . \quad (30)$$

It is however reasonable to assume that $P(x_u > S_u|I)$ is approximately 1 or 0, eg. when the standard deviation for the prediction is not extremely large. Hence we use the simpler rule

$$\langle L_u\rangle \approx \begin{cases} \langle L_{u1}\rangle & , \langle x_u\rangle > S_u \\ \langle L_{u2}\rangle & , \langle x_u\rangle \leq S_u \end{cases} , \quad (31)$$

where $\langle L_{u1}\rangle$ and $\langle L_{u2}\rangle$ are derived below with the results (42) and (43), and

$$\langle x_u\rangle = \sum_{r=1}^{R}\sum_{t=1}^{T}\rho_{urt}\langle c_{urt}\rangle \quad (32)$$

where

$$\langle c_{urt}\rangle = \int c_{urt}P(c_{urt}|I)dc_{urt} =$$

$$= \frac{1}{2}\left\{\hat{c}_{urt} + \hat{c}_{urt}\mathrm{erf}\left(\alpha_{urt}\hat{c}_{urt}\right) + \frac{1}{\sqrt{\pi}\alpha_{urt}}\left[\exp\left(-\alpha_{urt}^2\hat{c}_{urt}^2\right) - 1\right]\right\} \quad (33)$$

with

$$\alpha_{urt} = \frac{v-1}{\sqrt{2}\sigma_{urt}v} \quad . \quad (34)$$

The integral is straightforward and the proof is omitted.

Consider the calculation of $\langle L_{u1}\rangle$ which is the expectation with respect to $P(c_{urt}|I)$ of the corresponding case in (23). To distinguish between the expected loss with respect to $P(n_u|I)$ from (23) and the one currently under investigation we here assign the notation $\langle L_{u1}\rangle_{P(n_u|I)}$ for the former one.

Using the algebraic relation $x^{a+b} = x^a x^b$ we rewrite the expression for $x_u > S_u$ in (23) as

$$\langle L_{u1}\rangle_{P(n_u|I)} = \langle n_u\rangle\left(\frac{\langle n_u\rangle}{\langle n_u\rangle+1}\right)^{\sum_t^T\sum_r^R c_{urt}\rho_{urt}-S_u} = \langle n_u\rangle\left(\frac{\langle n_u\rangle}{\langle n_u\rangle+1}\right)^{-S_u}\prod_{t=1}^{T}\prod_{r=1}^{R}\left(\frac{\langle n_u\rangle}{\langle n_u\rangle+1}\right)^{c_{urt}\rho_{urt}} \quad .$$

Averaging over $P(c_{urt}|I)$ gives the expected loss contribution with respect to both $P(n_u|I)$ and $P(c_{urt}|I)$:

$$\langle L_{u1}\rangle = \langle n_u\rangle\left(\frac{\langle n_u\rangle}{\langle n_u\rangle+1}\right)^{-S_u} \times \prod_{t=1}^{T}\prod_{r=1}^{R}\int_{-\infty}^{\infty}P(c_{urt}|I)\left(\frac{\langle n_u\rangle}{\langle n_u\rangle+1}\right)^{c_{urt}\rho_{urt}}dc_{urt} \quad (35)$$

Inserting (26) into (35), the integral over c_{urt} contains three mutually exclusive intervals. We label the corresponding integrals I_1, I_2, and I_3. The first integral I_1

corresponding to the point $c_{urt} = \hat{c}_{urt}$ is simply

$$I_1 = \frac{1}{2}\left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right)^{\hat{c}_{urt}\rho_{urt}} . \quad (36)$$

The second integral, I_2, ranges from 0 to $\hat{c}_{urt}$. Using (28) we obtain

$$\begin{aligned} I_2 &= \int_0^{\hat{c}_{urt}} P_2(c_{urt}|I)\left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right)^{c_{urt}\rho_{urt}} dc_{urt} \quad (37)\\ &= \frac{1}{2} exp\left(\rho_{urt}\hat{c}_{urt}\log\left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right) + \rho_{urt}^2\gamma_{urt}^2\right)\times \\ &\times \left(\mathrm{erf}\left(\frac{(v-1)\hat{c}_{urt}}{v\sigma_{urt}\sqrt{2}} + \rho_{urt}\gamma_{urt}\right) - \mathrm{erf}\left(\rho_{urt}\gamma_{urt}\right)\right), \quad (38)\end{aligned}$$

where

$$\gamma_{urt} = \frac{\sigma_{urt} v}{(v-1)\sqrt{2}}\log\left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right) . \quad (39)$$

Finally, the third integral, I_3, represents the single point $c_{urt} = 0$ and using (29) we have:

$$\begin{aligned} I_3 &= \left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right)^{0\rho_{urt}}\left(\frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{(v-1)\hat{c}_{urt}}{v\sigma_{urt}\sqrt{2}}\right)\right) \quad (40)\\ &= \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{(v-1)\hat{c}_{urt}}{v\sigma_{urt}\sqrt{2}}\right) . \quad (41)\end{aligned}$$

Using I_1 from (36), I_2 from (38), and I_3 from (41) in (35) the expected loss contribution of user u with predicted capacities is, if $x_u > S_u$,

$$\langle L_{u1}\rangle = \langle n_u\rangle\left(\frac{\langle n_u\rangle}{\langle n_u\rangle + 1}\right)^{-S_u}\prod_{t=1}^{T}\prod_{r=1}^{R}(I_1 + I_2 + I_3) . \quad (42)$$

The second case in the expected loss contribution from user u (31) is simply

$$\begin{aligned}\langle L_{u2}\rangle &= \int P(c_{urt}|I)(S_u + \langle n_u\rangle - x_u)dc_{urt}\\ &= S_u + \langle n_u\rangle - \langle x_u\rangle \quad (43)\end{aligned}$$

The loss contributions in (42) and (43) are valid when predicted capacities can be modelled by a Gaussian distribution with known variance and expected value $\langle\hat{c}_{urt}\rangle = \bar{c}_{urt}$. It also requires that the obtained capacity decreases linearly when the predicted capacity $\hat{c}_{urt}$ is larger than the potential capacity $\bar{c}_{urt}$. It should however be emphasized that the linear decrease and the actual choice of v is a subjective choice, and not a property of the channel. The value of v depends on how sensitive the application is to departures from the desired BER. For low BER requirements, even a small prediction error leads to a substantial departure from the desired BER. For example, with uncoded M-QAM modulation[7], increasing from 4 bits to 5 bits per symbol at an SNR of 20 dB increases the BER by a factor of more than 200. (Use of coding increases the sensitivity.)

[7]Approximate BER formulas from [39] are used in these calculations.

Typically, in order to determine v we find the BER increase which means that the data must be retransmitted. We then determine the corresponding rate increase that would cause this BER discrepancy. If for instance M-QAM is used with a desired BER of 10^{-4}, and if a BER increase by a factor 100 would require that the data be retransmitted, then it can be found that $v \approx 1.5$ will be a good model. If a BER increase by a factor 10 would require retransmission, then $v \approx 1.2$. Typical values of v are thus in the range $1 < v < 2$. The linear decrease in c_{urt} for predictions larger than the potential capacity can be questioned, but clearly it satisfies the obvious requirement that the curve should be monotonic decreasing. Other alternatives would be to use either some concave of some convex decreasing function, but that could hardly make any substantial difference for the actual expected loss value unless the magnitude of the function's derivative would be very nearly zero for one interval and large for the remaining part. These cases will not be considered here, as they would rarely be encountered in practice.

The final expression (42) for $\langle x_u\rangle > S_u$ is rather complex and in the simulations of Section V-D we investigate whether the basic scheduler assuming perfect channel knowledge can be used with predicted values as an alternative to the more computationally burdensome minimization of (31). A simpler alternative to using (31) is however possible; note that we can approximately retain the desired property of lowering the predicted effective capacity when the uncertainty is high by using (23) with x_{urt} replaced by $\langle x_{urt}\rangle$ from (33). This approximation to (31) is exact if S_u is large compared to $\langle x_{urt}\rangle$.

C. Knowledge of Average Rates for Each Packet Size

We now consider the case where the sizes of incoming packets are known to the scheduler. The number of possible packet sizes is assumed small, for reasons we shall come back to in the derivations. Further, the expected number of incoming packets of each size in the time interval T is given. The effective capacities c_{urt} are here assumed known.

Let the packet sizes in the uth buffer, cf. Figure 1, belong to a set $\{k_u\}$ with K_u elements. Let m_{uk} be the number of packets of size k which are received in the uth buffer during the scheduling horizon T, with $\langle m_{uk}\rangle$ assumed known. In order to find a closed-form expression for the expected loss, we make a logic partitioning of each buffer u into K_u buffers. Hence, each user's buffer is split so that each packet size gets its own buffer. The remaining number of bits from the previous round, S_u, is also split into K_u partitions $S_u = \sum_k kS_{uk}$. Note however that this is only a logical separation for mathematical convenience.

Our new loss function is

$$L = \sum_{u=1}^{U}\sum_{k\in\{k_u\}} g\Big(km_{uk} + kS_{uk} - \frac{\sum_{t=1}^{T}\sum_{r=1}^{R} c_{urt}\rho_{urt}}{K_u}\Big), \quad (44)$$

where km_{uk} is the size (in bits) of the packet multiplied by the number of packets received by that size. It should be noted that the packet-enumerated loss function (44) is

equivalent to the bit-enumerated function (1). With the new loss function it is however easier to model knowledge of size-dependent packet-rates than when using (1).

For each user u we assign a probability distribution describing our knowledge of the future influxes m_{uk} corresponding to packets of size k. The probability assignment is analogous to (19):

$$P(m_{uk}|I) = \frac{1}{\langle m_{uk}\rangle + 1}\left(\frac{\langle m_{uk}\rangle}{\langle m_{uk}\rangle + 1}\right)^{m_{uk}} \quad , \tag{45}$$

and the resulting expected loss contribution of user u is

$$\langle L_u\rangle = \sum_{k\in\{k_u\}}\sum_{m_{uk}=0}^{\infty} P(m_{uk}|I)g\Big(km_{uk} + kS_{uk} - \frac{x_u}{K_u}\Big) \quad . \tag{46}$$

For each $k \in \{k_u\}$ we must separate between two possible cases, $\frac{x_u}{kK_u} > S_{uk}$ and $\frac{x_u}{kK_u} \leq S_{uk}$, which leads to different expressions. The derivation follows the procedure in Appendix B where (23) is derived. Consequently the total user contribution consists of the sum

$$\langle L_u\rangle = \sum_{k\in\{k_u\}} \langle L_{uk}\rangle \tag{47}$$

where

$$\langle L_{uk}\rangle = \begin{cases} k\langle m_{uk}\rangle\Big(\frac{\langle m_{uk}\rangle}{\langle m_{uk}\rangle+1}\Big)^{\frac{x_u}{kK_u}-S_{uk}} & , \ \frac{x_u}{kK_u} > S_{uk} \\ k\langle m_{uk}\rangle + kS_{uk} - \frac{x_u}{K_u} & , \ \frac{x_u}{kK_u} \leq S_u \end{cases} \quad . \tag{48}$$

It should be noted that if there is a wide variety of packet sizes, i.e. if K_u is large, then the expression above would consist of too many terms for it to be tractable in actual calculations. We should then instead assign a probability density for n_u, the number of incoming bits in each buffer. This is possible (see [1] for a similar derivation) and results in a Gaussian approximation. The derivation is rather lengthy, and it is not presented here due to space considerations.

V. Comments and Simulations

By using prior probability distributions with maximum entropy subject to our information constraints, we avoid assumptions concerning the "underlying" long-run behavior of the sources. The use of the maximum entropy distribution is motivated because it is the distribution which can arise in the greatest number of ways when the outcomes are constrained to agree with the given information [22].

Other reasonable approaches to modelling the influxes include using more information in the initial probability assignments, and adapting the distributions according to incoming data using Bayes' theorem. For instance, if we have knowledge of correlations over time or among different user streams, then we can use this information in the maximum entropy formalism to obtain prior distributions of lower entropy than using the mean values only. If such correlations are known to exist but their absolute values are unknown *a priori*, then the initial probability distribution should be updated recursively according to Bayes' theorem as observations of the data streams become available. More research needs to be directed towards finding methods that can infer patterns in on-going data streams and adapt posterior distributions with low complexity. A step in this direction has been taken in very recent work [40], but more work is needed for the specific case of individual data streams.

A. On the optimality of time division multiple access (TDMA)

Previous work [41] claims that time division is an optimal scheduling policy in CDMA on the grounds that it minimizes the received power levels from other users. However, in CDMA systems, the bad effects of interference are alleviated by well-designed codes. The interfering users' signal levels are not necessarily harmful to the detection performance of the desired user and thus we cannot conclude that it is always optimal to use time division.

In spite of this one might conjecture that, would the buffers never be emptied, it might be optimal to use time division also when interference does not affect receiver performance. This conjecture was proven to be true in the deterministic case in the sense of maximizing the sum-of-rates capacity of an uplink in a multiuser single-cell scenario by Knopp and Humblet [3] when the time-varying fading channels were perfectly tracked and known at the transmitters. In general, however, neither source rates nor channels are perfectly known and buffers may be emptied. Hence, time division is not always an appropriate choice. To see this, consider the problem of scheduling one channel one time slot at a time, ie. $R = 1$, $T = 1$. It can be observed from the expected loss expression (23) that if the buffer contents of the user with the highest effective capacity c_{ut} satisfies $S_u \geq c_{ut}$, then the minimum loss is obtained by transmitting exclusively to that user. If this condition is not met, then we cannot conclude that exclusive transmission is optimal in the sense of maximizing expected throughput.

Example V.1: Consider the problem of assigning bandwidth across two users using one channel and one time slot, i.e. $U = 2$, $R = 1$, $T = 1$. Assume that the users have $S_1 = S_2 = 10$ bits in stock and their expected influx for the next time slot is $\langle n_1\rangle = \langle n_2\rangle = 10$. Assume knowledge of the effective capacities, $c_1 = 17$ and $c_2 = 20$.

Figure 5 plots the total expected buffer contents using (23) as a function of $\rho_1 = 1-\rho_2$. The optimum assignment is to split the bandwidth almost equally among the users. Even though the user with the highest capacity seems to have a large probability for being able to transmit 20 bits (since $S_2 + \langle n_2\rangle = 20$) the uncertainty is still considerable and the best decision is to refrain from exclusive transmission. The probability that $n_2 = 0$ is large, and we can only be certain about transmitting 10 bits (the number of bits already in stock) to user 2. Therefore, it would be unnecessarily risky to let user 2 obtain all bandwidth when we know for certain that it can be used to reduce the buffer

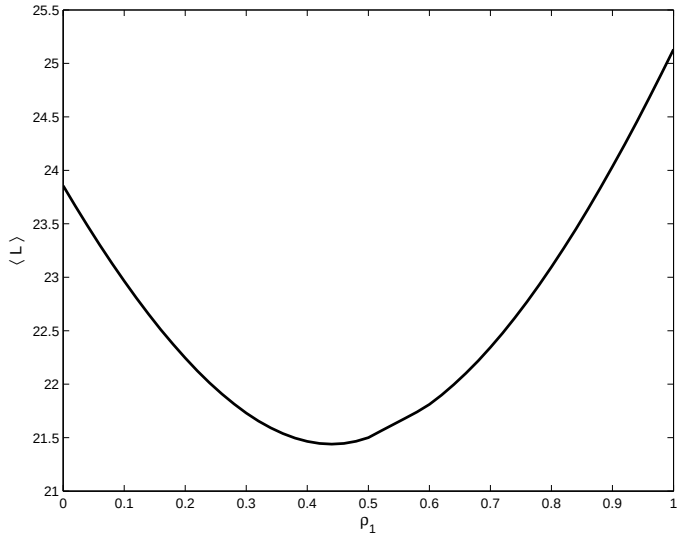


Fig. 5. The expected loss using (23) as a function of $\rho_1 = 1 - \rho_2$ for the scenario in Example V.1.

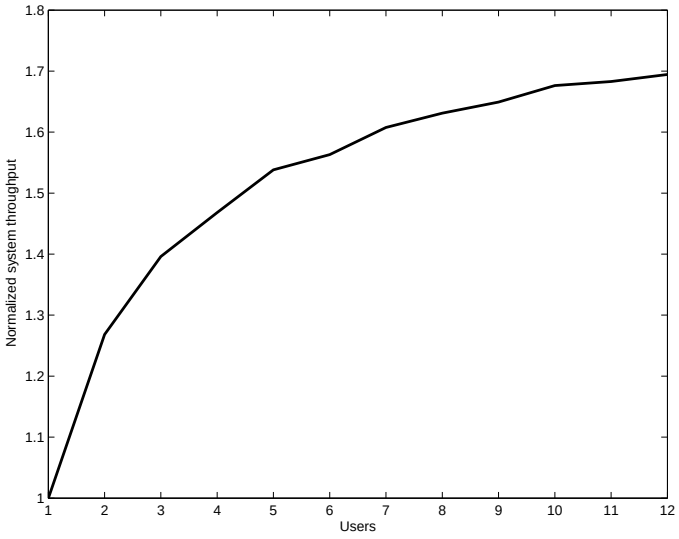


Fig. 6. The total downlink throughput obtained in a system employing the basic scheduler increases with the number of users. Each user experienced independent Rayleigh fading on the time scale of slots, with an average SNR of 10 dB.

levels of user 1. □

If the scheduler uses a longer time horizon, the minimum loss is obtained with exclusive allocations for each time slot if for every time slot the user with maximum capacity at that time fulfills the criterion $S_u \geq c_{ut}$. If there at any time slot is some user with maximum effective capacity having less data to send than the channel allows, no general conclusion about the optimality of exclusive transmission at any time slot can be drawn. We may conjecture that the scheduler will indeed use exclusive assignments also in many cases that are not covered by the general conditions for optimality; the loss expression does however not give any simple criterion for this to be the optimal choice in general.

Further, for the conjecture to be true, the transmission resources (consisting of antennas, codes, modulation format, etc.) must be such that there is no additional advantage of letting two users transmit at the same time. For instance, some resources might not be mutually exclusive, i.e. two users may utilize them fully at the same time. The model used throughout this paper does not consider such resources.

B. Multiuser diversity gain

In this section we investigate how the capacity of a system increases with the number of users when utilizing multiuser diversity.

In Figure 6 the sum throughput is plotted as a function of the number of users in a simulated system. The results were obtained using the basic scheduler with perfect channel knowledge using (23) in a scenario with two parallel independently fading channels ($R = 2$). Each user experienced independent Rayleigh fading on the time scale of slots, and the effective capacity was modelled as the integer nearest below the Shannon capacity for a channel disturbed by additive white Gaussian noise only[8],

$$c_{urt} = \log_2 (1 + \gamma_{urt}) \ , \tag{49}$$

where γ_{urt} denotes the SNR at the receiver. Assuming one-tap Rayleigh fading, γ_{urt} is exponentially distributed. The average SNR was set to 10 dB, and the source rates were set so that the transmission buffers were never emptied.

Define the multiuser diversity gain, or scheduling gain, α, as the ratio between the obtained total throughput, x, and the throughput that would have been obtained by simple round-robin scheduling, $x^{(RR)}$,

$$\alpha = \frac{x}{x^{(RR)}} \ . \tag{50}$$

Figure 6 then describes the scheduling gain of the simulated scenario, since round-robin scheduling gives a sum throughput equal to the average effective capacity for any one of the users.

C. Comparison with proportional fair scheduling

In a new set of simulations the proportional fair scheduler (see eg. [8]) was compared to the basic scheduler from Section IV-A with knowledge of effective capacities (using (23)). Both these schedulers use knowledge of the channel to guide their decisions. The proportional fair scheduler does however not consider the effects of source rates and hence the possibility of empty buffers. Implicitly it assumes that there is always data to send.

The proportional fair scheduler works as follows. The data rates that the users' channels can support at each time slot t (the effective capacity) is known to the scheduler. The scheduler then keeps track of the average throughput

[8]The model used here would in reality require perfect channel adaptation and a continuum of modulation levels and coding rates.

TABLE I

Parameters for the comparison of proportional fair scheduling with the maximum entropy scheduler for known channels. Average inflows per time slot, $\frac{\langle n_i \rangle}{T}$, average SNR (dB) at the receiver, γ_u, and the corresponding average effective channel capacity (number of bits per time slot), $\langle c_{urt} \rangle$.

	$\frac{\langle n_i \rangle}{T}$	$\gamma_u(dB)$	$\langle c_{urt} \rangle$
User 1	2	10	2.9
User 2	6	13	3.7
User 3	1.5	13	3.7

TABLE II

Results for the comparison of proportional fair scheduling with the maximum entropy scheduler for known channels. The average number of bits remaining in the buffers after 60 time slots are listed in columns 1 and 2 for the proportional fair scheduler (PF) and the scheduler with known c_{urt} proposed here (ME). The last two columns display average total throughput in bits.

	S_{60}(PF)	S_{60}(ME)	Tp(PF)	Tp(ME)
User 1	2	11	117	108
User 2	170	35	191	326
User 3	0	4	92	88
Total	172 bits	50 bits	400 bits	522 bits

$T_u(r,t)$ of each user u in a past window of length t_c. On each channel r and time slot t, the scheduler transmits exclusively to the user with the largest $\frac{c_{urt}}{T_u(r,t)}$. The parameter t_c is used as a forgetting factor in the calculation of the windowed average throughput. It is used as a means of obtaining fairness, by giving a user access to a channel when its effective capacity is high relative to its own average throughput over the time scale t_c. In [8], a single base station is considered. Here, we adapt the proportional fair scheduler to multiple parallel channels simply by treating an additional channel as additional time slots. In other words, if we are to assign two channels and three time slots, the scheduler works exactly as if it were to schedule one channel and six time slots. After each single assignment, the average throughput $T_u(i)$ (where i indexes assignments regardless of whether it describes time slot or channel) is recalculated according to [43]:

$$T_u(i) = (1 - \frac{1}{t_c})T_u(i-1) + \frac{1}{t_c}c_{u,i-1}\delta(u - u^*) \quad , \qquad (51)$$

where $\delta(u - u^*) = 1$ if user u was the transmitting user u^* in the most recent assignment, otherwise, $\delta(u - u^*) = 0$.

The schedulers were run on the same data sets, with source rates n_{ut} drawn from a Poisson random number generator[9], and effective capacities generated from the rate expression (49) using an exponential pdf for the SNR. The parameters used are listed in Table I. The forgetting factor for the proportional fair scheduler was set to $t_c = 7$.

The simulated scenario consisted of two parallel independently fading channels ($R = 2$) and three users ($U = 3$). The scheduling horizon was $T = 3$ time slots, and the schedulers were run for a total of 60 time slots. The results listed in Table II are averages from 100 realizations. The table reports average throughput and average buffer levels after the 60 time slots.

The results show that in this scenario the total throughput increases by approximately 30% using (23) compared with using the proportional fair scheduler. In particular, the throughput of user 2 is severely degraded when buffer contents are neglected. In terms of buffer levels it is clear that the second user's buffer would overflow, causing further throughput degradation and increasing delays due to the invoking of higher-layer mechanisms such as decreasing transmission rates or retransmissions.

Comparing the results for users 2 and 3, having equal channel statistics, we see that the throughput ratio of the two users is identical to the ratio of their average inflows when using maximum entropy scheduling. If the inflows are taken to reflect each user's service requirements, then this means that fairness is obtained without any explicit fairness constraint on the policy. On the other hand, a user with very low average SNR and small channel variability would obviously risk starvation with the proposed scheduler.

It can be noted that a maximum SNR scheduler (which is a special case of the proportional fair scheduler when all users have independent but identical channel statistics) could approach the performance of the maximum entropy scheduler were the transmission buffers constantly flooded with data. A more important observation is that this case is normally prevented from occurring in a real system due to rate-control mechanisms such as provided by TCP. Schedulers should therefore always take buffer contents into account. The additional use of *source rate diversity* further increases the performance of the maximum entropy scheduler.

Another interesting result from this simulation can be observed by studying the throughput obtained for the second user, 326 bits. Instead of trying to use multiuser diversity to our advantage we could split the available bandwidth into three equal parts, and always transmit to all users. Instead of 326 bits, user 2 would then obtain a total throughput of $\frac{3.7}{3} \times 2 \times 60 = 148$ bits. Thus, the individual throughput increases by 120% when using the fluctuating channels and arrival rates as sources of diversity. The proportional fair scheduler only achieves an increase of 29% since it does not take the varying arrival rates into account. Evidently, there are substantial benefits associated with taking advantage of the fact that, on average, the other users' arrival rates are lower than their effective capacities. Neglecting this source of diversity results in decreased individual and total throughput.

D. Results for different amounts of channel uncertainty

Having established that channel information and taking arrival rates into account are critical issues, two questions

[9]This choice is admittedly somewhat arbitrary. For a discussion of the problems involved in modelling and simulating individual Internet sources see [35].

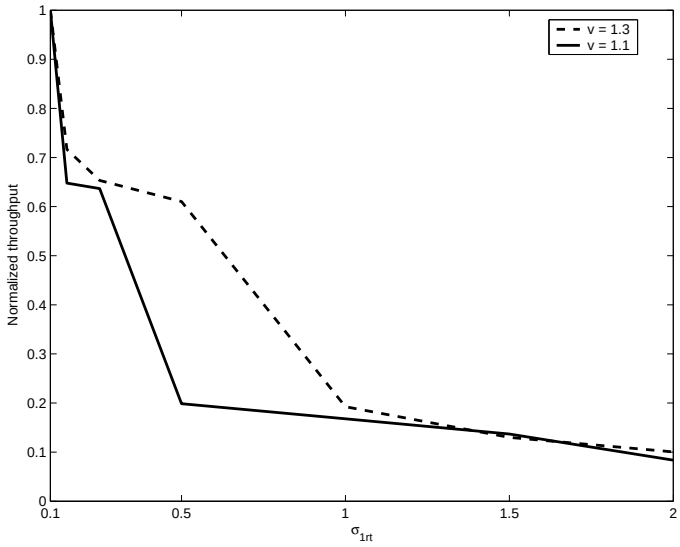


Fig. 7. The normalized throughput (1 corresponding to the throughput of user one if $\sigma_{1rt} = 0.1$) for user one as a function of σ_{1rt}. All users had the same average source rates and potential effective capacities ($\langle \bar{c} \rangle \approx 2.9$) (cf. Section V-D). The two curves correspond to different values of the BER sensitivity v.

naturally arise:
1. How does the accuracy of channel predictions affect individual and total throughput?
2. Do we need to use the more complex scheduler when using inaccurate channel predictions or can we equally well use the simpler one, assuming perfect channel knowledge?

To answer the first question, we study the throughput degradation of a user as a function of increasing prediction inaccuracy. The simulation setup consists of scheduling six users according to (31), with two independently fading and non-interfering channels, $R = 2$, and a scheduling horizon of $T = 3$ time slots. All users have an average SNR of 10 dB, and the Rayleigh fading model from Section V-B is used with the effective capacity described by (49). (The average potential effective capacity is thus approximately 2.9 bits.) The buffer influxes are large compared to the effective capacities. All users except the first one have nearly perfect prediction, $\sigma_{urt} = 0.1$. During a simulation run for 60 time slots, user one's prediction accuracy was held at a constant value. The simulation was then repeated for a range of increasing prediction inaccuracies $\sigma_{1rt} = 0.1 \ldots 3.5$. Figure 7 shows the throughput of user one for two different BER sensitivities, $v = 1.3$ and $v = 1.1$. We see that the throughput degrades very quickly for decreasing prediction accuracy. Already at $\sigma_{1rt} = 0.15$ the throughput has degraded to roughly 60% of what a user with $\sigma_{1rt} = 0.1$ obtains. The reason is that there is almost always another user with equally high predicted capacity, but with higher accuracy, thereby leaving user one at a disadvantage since a larger uncertainty σ_{urt} results in lower expected effective capacity (33).

In terms of an individual user's performance at any specific time slot, therefore, an important property of the predictor is that its accuracy should be comparable to that of the other users. On the level of system throughput, however, since the expected throughput $\langle x_{urt} \rangle$ decreases with prediction inaccuracy, the total throughput necessarily decreases too if the accuracy is equal among users. But if the accuracy varies independently among users, it is likely that there is at least one user with both high SNR and good accuracy. From a system throughput perspective, therefore, prediction accuracy should preferably vary across users. As long as each user has *on average* similar prediction accuracy as other users, this is indeed desirable for individual users as well. Furthermore, prediction accuracy in the high-SNR region is more important than for low SNR, since a user will only be scheduled for transmission in the former case.

Addressing the second question, a simulation setup was run comparing (31) with the basic scheduler using (23) but employing the predicted values of the effective capacity, $\hat{c}_{urt}$, instead of the true values. Here, all parameters except the prediction accuracies were the same as for the previous simulation; the prediction accuracies varied independently among users, channels and time slots according to a uniform probability distribution, $\sigma_{urt} \in [0, 3]$.

The sum throughput using (31) was 17% higher than when using (23) for $v = 1.1$, and 16% higher for $v = 1.3$. The performance difference between the two schedulers[10] can be interpreted as a third diversity dimension; in addition to multiuser diversity and source rate diversity, the *prediction accuracy diversity* allows the full Bayesian solution (31) to pick a user with both high effective channel capacity and high prediction accuracy. This implies that the more complex scheduler should be used in situations where different users have different prediction accuracies, for instance due to different user velocities.

VI. Conclusions

In this paper a problem of optimizing resource assignments in the presence of uncertainty was considered for applications in mobile communications. The problem was formulated as a minimization of the expected total buffer contents, given by the general expression (4), a sum of contributions from each user. It was noted that the framework is compatible with user priorities represented by deterministic functions describing an equivalent cost per bit.

Each user's contribution to the total expected loss was calculated for three different cases, each representing a typical state of knowledge at the scheduler. With knowledge of effective capacities and of average influxes, the expected loss contribution was found in (23). Using knowledge of the accuracy of capacity predictions, a Gaussian distribution was assigned for the predicted capacities. It was noted that the obtained capacity is a function of the prediction, and the resulting probability distribution for the effective capacities was derived for the case when too large predictions result in a linear decrease of obtained capacity. The

[10]Notice that if all users would have had the same prediction accuracy (this is unlikely, since different users move at different velocities), then there would not have been any performance difference between the two schedulers, since using (31) would merely reduce all users' expected capacity by a nearly equal amount.

consequent expected loss contribution was found in (31). In a packet data system with knowledge of packet sizes, effective capacities, and average influxes for each packet size, the resulting expected loss contribution was described by (46).

A substantial increase in throughput due to multiuser diversity gain from maximum entropy scheduling was demonstrated in simulations. A comparison of maximum entropy scheduling with the proportional fair scheduler showed that the maximum entropy scheduler achieved higher throughput by also utilizing source rate diversity. Further simulations demonstrated that in order to obtain high throughput the scheduler needs to have accurate channel knowledge. Degradation of channel prediction accuracy for one user inevitably led to reduced throughput for that user as described by Figure 7. Including knowledge of prediction accuracy into the criterion resulted in improved system performance compared to using the basic criterion with predicted capacities instead of the true values. The performance difference was a consequence of exploiting the prediction accuracy diversity. The larger the variations in channel prediction accuracy and the more users in the system, the larger the resulting gain of using the full Bayesian solution (31).

The Bayesian solution thus prioritizes users with well-determined high-rate channels, and with data to send. In the limit, as the number of users tends to infinity and the prediction accuracies vary independently over the users, the full Bayesian solution would approach the throughput of the scheduler with perfect channel knowledge, since then there would almost always exist a user having maximum effective capacity with negligible uncertainty.

Observe also that the expected loss expressions could be used in other types of schedulers as well. For instance, with strict delay requirements, a simple and effective scheme for exclusive one-slot scheduling would be to transmit to the user u who yields the largest total loss decrease, $\langle L(\rho_{ur} = 0)\rangle - \langle L(\rho_{ur} = 1)\rangle$ (which is the best exclusive scheduling policy in the sense of minimizing expected loss). Then at the next time slot, the remaining $U-1$ users would compete similarly. For each time slot, the set of competing users is reduced, and after U time slots, the process repeats. The maximum delay for any user would then be $2U-1$ time slots. This type of scheduling policy with reduced channel feedback is investigated further in [44].

In conclusion it should be pointed out that, although the framework was formulated in a communication theoretic setting, the rationale can be employed in other forms of resource optimization problems where the demand, n_u, is incompletely known. The case of incompletely known supply, c_{urt} corresponding to the solution laid out in Section IV-B, would however require a different supply distribution than here. This is in principle straightforward; given any testable information regarding the actual supply mechanisms, find the $P(c_{urt}|I)$ that maximizes the corresponding entropy. Given that model, the solution that maximizes the number of satisfied orders is again given by (4).

Appendices

A.

The maximum entropy distribution is found using the Lagrange method. Using the constraints (10) we form the functional

$$H(P) = -\sum_{i=1}^{n} P_i \log P_i + \sum_{k=1}^{m} \lambda_k \left(F_k - \sum_{i=1}^{n} P_i f_k(x_i) \right) \tag{52}$$

and differentiate with respect to P_i:

$$\frac{\partial H(P)}{\partial P_i} = -\log P_i - 1 - \sum_{k=1}^{m} \lambda_k f_k(x_i) \ . \tag{53}$$

Setting this equal to zero we have the general form of the entropy-maximizing probability mass:

$$P_i = \exp\big[-1 - \sum_{k=1}^{m} \lambda_k f_k(x_i)\big] \ . \tag{54}$$

However we have not yet included the constraint that $\sum_{i=1}^{n} P_i = 1$. This is just a normalization, and we obtain:

$$P_i = \frac{1}{\sum_{i=1}^{n} \exp\big[-\sum_{k=1}^{m} \lambda_k f_k(x_i)\big]} \exp\big[-\sum_{k=1}^{m} \lambda_k f_k(x_i)\big] \ . \tag{55}$$

The Lagrange multipliers λ_i are chosen so that the constraints (10) are satisfied.

This procedure is formulated in a compact form by introducing the partition function (12) and rewriting (55) as

$$P_i = \frac{1}{Z(\lambda_1, \ldots, \lambda_m)} \exp\big[-\sum_{k=1}^{m} \lambda_k f_k(x_i)\big] \ . \tag{56}$$

In order to find the Lagrange multipliers satisfying the constraints (10) we notice that differentiating $\log Z$ with respect to each λ_k gives:

$$\begin{aligned} \frac{\partial}{\partial \lambda_k} \log Z &= \frac{1}{Z(\lambda_1, \ldots, \lambda_m)} \sum_{i=1}^{n} \big(-f_k(x_i) \times \\ &\times \exp\big[-\lambda_1 f_1(x_i) - \ldots - \lambda_m f_m(x_i)\big]\big) \\ &= -\sum_{i=1}^{n} P_i f_k(x_i) \ , \end{aligned} \tag{57}$$

which is the formulation of the constraints (10).

Thus the constraints (10) are satisfied by choosing the Lagrange multipliers so that

$$F_k = -\frac{\partial}{\partial \lambda_k} \log Z \ . \tag{58}$$

B.

In Section IV-A, in the derivation of the expected loss contribution assuming knowledge of effective capacities and average source rates, we need to evaluate the summation

over n_u in (22). Using the probability assignment (20) for the influxes we obtain:

$$\sum_{n_u=0}^{\infty} P(n_u|I)g(S_u+n_u-x_u) = \begin{cases} \langle L_\dagger \rangle & , \ x_u > S_u \\ S_u + \langle n_u \rangle - x_u & , \ x_u \leq S_u \end{cases} . \tag{59}$$

where

$$\begin{aligned}
\langle L_\dagger \rangle &= \sum_{n_u=0}^{\infty} P(n_u|I)(S_u+n_u-x_u) \\
&\quad - \sum_{n_u=0}^{x_u-S_u} P(n_u|I)(S_u+n_u-x_u) \\
&= \sum_{n_u=0}^{\infty} \frac{1}{\langle n_u \rangle + 1}\Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u}(S_u+n_u-x_u) \\
&\quad - \sum_{n_u=0}^{x_u-S_u} \frac{1}{\langle n_u \rangle + 1}\Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u}(S_u+n_u-x_u) \\
&= \sum_{n_u=0}^{\infty} \frac{1}{\langle n_u \rangle + 1}\Big[\Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u} n_u + && (60) \\
&\quad + \Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u}(S_u - x_u)\Big] && (61) \\
&\quad - \sum_{n_u=0}^{x_u-S_u} \frac{1}{\langle n_u \rangle + 1}\Big[\Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u} n_u + && (62) \\
&\quad + \Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{n_u}(S_u - x_u)\Big] && (63) \\
&= \langle n_u \rangle + S_u - x_u && (64) \\
&\quad - (S_u - x_u)\Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{x_u-S_u+1} && (65) \\
&\quad - \langle n_u \rangle \Big(1 - \Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{x_u-S_u}\Big) && (66) \\
&\quad - (S_u - x_u)\Big(1 - \Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{x_u-S_u+1}\Big) && (67) \\
&= \langle n_u \rangle \Big(\frac{\langle n_u \rangle}{\langle n_u \rangle + 1}\Big)^{x_u-S_u} . && (68)
\end{aligned}$$

The infinite progression in lines (60) and (61) are standard sums which can be found in [45] (eqns. 0.231.2 and 0.231.1). They correspond to the solution (64). The finite sum in lines (62) and (63) can also be found in [45] (eqns. 0.113 and 0.112). The arithmetico-geometric progression (62) corresponds to the solution spanning lines (65) and (66), while the geometric series (63) corresponds to the solution (67).

C.

Here we derive the expected loss contribution for known time-varying influx averages, assuming perfect knowledge of the effective capacities. The probabilities for n_{ut} for different times t factor according to the maximum entropy principle and thus we can rewrite the expected loss contribution as a product of independent terms. As in (23) we need to separate between the cases $x_u > S_u$ and $x_u \leq S_u$. It follows immediately from the derivation of (23) in Appendix B that for $x_u \leq S_u$ the loss contribution for user u is

$$\begin{aligned}
\langle L_u \rangle &= S_u + \textstyle\sum_{t=1}^{T} \langle n_{ut} \rangle - x_u \\
&= S_u + \langle n_u \rangle - x_u \quad , \ x_u \leq S_u .
\end{aligned} \tag{69}$$

Consider the calculation of $\langle L_u \rangle$ in the case $x_u > S_u$. For reasons we shall come back to in the derivation we need to reorder the $\langle n_{ut} \rangle$ by decreasing size. Thus, we replace the time indexes t by size indexes k, where larger k corresponds to smaller size. We start by deriving the average loss with respect to $P(n_{u1}|I)$, for given smaller influxes $n_{u2}, n_{u3}, \ldots$, which we denote by $\langle L_u \rangle_{P(n_{u1}|I)}$. By substituting $S_u + \sum_{k=2}^{T} n_{uk}$ for S_u in the derivation of (23) in Appendix B it follows directly that

$$\begin{aligned}
\langle L_u \rangle_{P(n_{u1}|I)} &= \langle n_{u1} \rangle \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{x_u - S_u - \sum_{k=2}^{T} n_{uk}} \\
&= \langle n_{u1} \rangle \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{x_u - S_u} \prod_{k=2}^{T} \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{-n_{uk}} .
\end{aligned} \tag{70}$$

This means that the expected loss averaged over the influxes at the remaining times, $n_{u2}, \ldots,$ becomes

$$\begin{aligned}
\langle L_u \rangle &= \langle n_{u1} \rangle \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{x_u - S_u} \\
&\quad \times \prod_{k=2}^{T} \sum_{n_{uk}=0}^{\infty} P(n_{uk}|I) \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{-n_{uk}} .
\end{aligned} \tag{71}$$

The sum over n_{uk} in (71) is, by using (20), given by

$$\begin{aligned}
&\sum_{n_{uk}=0}^{\infty} \frac{1}{\langle n_{uk} \rangle + 1}\Big(\frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1}\Big)^{n_{uk}} \Big(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\Big)^{-n_{uk}} \\
&\quad = \sum_{n_{uk}=0}^{\infty} \frac{1}{\langle n_{uk} \rangle + 1}\left(\frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1}\frac{\langle n_{u1} \rangle + 1}{\langle n_{u1} \rangle}\right)^{n_{uk}} && (72) \\
&\quad = \frac{1}{\langle n_{uk} \rangle + 1}\left(\frac{1}{1 - \frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1}\frac{\langle n_{u1} \rangle + 1}{\langle n_{u1} \rangle}}\right) . && (73)
\end{aligned}$$

In the last equality the reordering of $\langle n_{uk} \rangle$ by decreasing size is needed to ensure convergence of the geometric series (72) (eqn. 0.231.1 in [45]) , which requires $\frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1}\frac{\langle n_{u1} \rangle + 1}{\langle n_{u1} \rangle} < 1$. The average loss is then

$$\begin{aligned}
\langle L_u \rangle &= \langle n_{u1} \rangle \left(\frac{\langle n_{u1} \rangle}{\langle n_{u1} \rangle + 1}\right)^{x_u - S_u} \\
&\quad \times \prod_{k=2}^{T} \frac{1}{\langle n_{uk} \rangle + 1}\left(\frac{1}{1 - \frac{\langle n_{uk} \rangle}{\langle n_{uk} \rangle + 1}\frac{\langle n_{u1} \rangle + 1}{\langle n_{u1} \rangle}}\right) .
\end{aligned} \tag{74}$$

D.

In Section IV-B the probability for the obtained effective capacity c_{urt} given a prediction is needed in order to calculate the expected loss. We derive the probability for each of the three cases (cf. Figure 3) and then add the resulting distributions to obtain the total probability distribution.

1. When $\hat{c}_{urt} \leq \bar{c}_{urt}$ the obtained capacity is $c_{urt} = \hat{c}_{urt}$. Because the distribution for the predicted capacity is symmetric and centered at the potential capacity $\bar{c}_{urt}$ we have

$$P_1(c_{urt}|I) = \frac{1}{2}\delta(c_{urt} - \hat{c}_{urt}) \tag{75}$$

where δ is the Dirac delta.
2. In the second interval, $\bar{c}_{urt} \leq \hat{c}_{urt} \leq c^*_{urt}$, we use the aforementioned linearly decreasing function in describing the obtained capacity:

$$c_{urt} = -\frac{1}{v-1}\hat{c}_{urt} + \frac{v}{v-1}\bar{c}_{urt} \ . \tag{76}$$

Leaning on previous remarks we model the potential capacity as a Gaussian distribution according to $\bar{c}_{urt} \sim \mathcal{N}(\hat{c}_{urt}, \sigma^2_{urt})$. Using the result

$$x \sim \mathcal{N}(m \ , \ \sigma^2) \Rightarrow ax + b \sim \mathcal{N}(am + b \ , \ a^2\sigma^2) \tag{77}$$

and the relation (76) it is concluded that

$$\begin{aligned} c_{urt} &\sim \mathcal{N}\left(-\frac{1}{v-1}\hat{c}_{urt} + \frac{v}{v-1}\hat{c}_{urt} \ , \left(\frac{v\sigma_{urt}}{v-1}\right)^2\right) \\ &= \mathcal{N}\left(\hat{c}_{urt} \ , \left(\frac{v\sigma_{urt}}{v-1}\right)^2\right) \ . \end{aligned} \tag{78}$$

Notice that this distribution is attained only for the interval $0 \leq c_{urt} \leq \hat{c}_{urt}$.
3. In the third interval, $\hat{c}_{urt} \geq v\bar{c}_{urt}$ or equivalently $-\infty \leq \bar{c}_{urt} \leq \hat{c}_{urt}/v$, the obtained capacity is zero. The probability for this is

$$\begin{aligned} P_3(c_{urt}|I) &= \delta(c_{urt}) \int_{-\infty}^{\hat{c}_{urt}/v} P(\bar{c}_{urt}|I) d\bar{c}_{urt} \\ &= \delta(c_{urt}) \int_{-\infty}^{\hat{c}_{urt}/v} \frac{1}{\sqrt{2\pi\sigma^2_{urt}}} \exp\Big[-\frac{1}{2\sigma^2_{urt}}(\bar{c}_{urt} - \hat{c}_{urt})^2\Big] d\bar{c}_{urt} \\ &= \delta(c_{urt}) \left(\frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{(v-1)\hat{c}_{urt}}{v\sigma_{urt}\sqrt{2}}\right)\right) \ , \end{aligned} \tag{79}$$

where erf(x) is the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \ . \tag{80}$$

Gaussian integrals like the previous one are solved by combining eqns. 3.322.1, 3.322.2, and 3.323.2 in [45].

References

[1] E. T. Jaynes, "New engineering applications of information theory", *Engineering Uses of Random Function Theory and Probability*, Bogdanoff and Kozin (eds.), Wiley, New York, pp 163-203, 1963.

[2] E. T. Jaynes, *Probability Theory – The Logic of Science*, Cambridge University Press, April 2003.

[3] R. Knopp, P.A. Humblet, "Information capacity and power control in single-cell multiuser communications", Proc. IEEE ICC 95, June 1995.

[4] Raymond Knopp, *Coding and Multiple-Access over Fading Channels*, D.Sc. Thesis, Swiss Federal Institute of Technology (Lausanne), Dept. of Electrical Engineering, 1997.

[5] Yaxin Cao, Victor O. K. Li, "Scheduling algorithms in broadband wireless networks", Proceedings of the IEEE, Vol. 89, No. 1, Jan 2001.

[6] Nilo Casimiro Ericsson, *On Scheduling and Adaptive Modulation in Wireless Communications*, Licentiate Thesis, Signals & Systems Group, Uppsala University, June 2001.

[7] Nilo Casimiro Ericsson, Sorour Falahati, Anders Ahlén, Arne Svensson, "Hybrid type-II ARQ/AMS supported by channel predictive scheduling in a multi-user scenario", Proc. IEEE VTC Fall 2000, Sep 24-28 2000.

[8] Pramod Viswanath, David N.C. Tse and Rajiv Laroia, "Opportunistic beamforming using dumb antennas", IEEE Transactions on Information Theory, Vol. 48, No. 6, June 2002.

[9] D. N. Tse, "Optimal power allocation over parallel Gaussian channels", Proc. IEEE ISIT, June 1997.

[10] Leandros Tassiulas and Anthony Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks", IEEE Transactions on Automatic Control, Vol. 37, No. 12, Dec. 1992.

[11] Leandros Tassiulas and Anthony Ephremides, "Allocation of a single server to a set of parallel queues with time dependent demands", Proc. IEEE ISIT, June 1991.

[12] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, "Providing quality of service over a shared wireless link", IEEE Communications Magazine, Feb. 2001.

[13] Farrokh Rashid-Farrokhi, Leandros Tassiulas, K. J. Ray Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays", IEEE Transactions on Communications, Vol. 46, No. 10, Oct 1998.

[14] Roy D. Yates and Ching-Yao Huang "Integrated Power Control and Base Station Assignment" IEEE Transactions on Vehicular Technology, Vol. 44, No. 3, Aug 1995.

[15] Mats Bengtsson, "Jointly optimal downlink beamforming and base station assignment", Proc. ICASSP - 2001, May 2001.

[16] G. Caire and S. Shamai Shitz, "On the achievable throughput of a multiantenna Gaussian broadcast channel", IEEE Transactions on Information Theory, Vol. 49, No. 7, pp. 1691-1706, July 2003.

[17] N. Jindal, S. Vishwanath, A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels", IEEE Transactions on Information Theory, Vol. 50, No. 5, pp 768-783, May 2004.

[18] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality", IEEE Transactions on Information Theory, Vol. 49, No. 8, pp. 1912-1921, Aug 2003.

[19] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial channel state information", Submitted to IEEE Transactions on Information Theory, June 2003.

[20] E. T. Jaynes, "Information theory and statistical mechanics", The Physical Review, Vol. 106, No. 4, pp. 620-630, May 15 1957.

[21] E. T. Jaynes, "Information theory and statistical mechanics II", The Physical Review, Vol. 108, No. 2, pp. 171-190, Oct 15 1957.

[22] E. T. Jaynes, "On the rationale of maximum-entropy methods", Proceedings of the IEEE, Vol. 70, No. 9, pp. 939-952, Sept 1982.

[23] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data", Nature, Vol. 272, p. 686, 1978.

[24] G. J. Daniell and S. F. Gull, "The maximum entropy algorithm applied to image enhancement", Proceedings of the IEEE, Vol. 5, No. 127, p. 170, 1980.

[25] J. P. Burg, "Maximum entropy spectral analysis", Proc. 37th Meet. Soc. Exploration Geophysicists, 1967; Stanford Thesis 1975.

[26] P. W. Buchen and M. Kelly, "The maximum entropy distribution of an asset inferred from option prices", The Journal of Financial and Quantitative Analysis, Vol. 31, No. 1, pp. 143-159, Mar 1996.

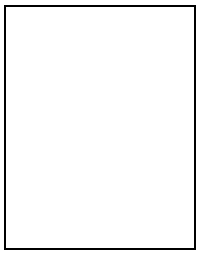
[27] R. Rosenfeld, "A maximum entropy approach to adaptive statistical language modelling", Computer Speech and Language, Vol. 10, Issue 3, pp. 187-228, 1996.

[28] W. T. Grandy Jr., "Principle of maximum entropy and irreversible processes", Physics Reports, Vol. 62, Issue 3, pp. 175-266, July 1980.

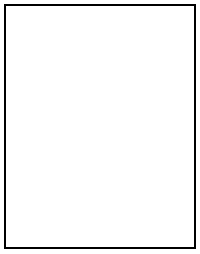
[29] J. L. Gruver, J. Aliaga, H. A. Cerdeira, and A. N. Proto, "Nontrivial dynamics induced by a Jaynes-Cummings Hamiltonian", Physics Letters A, Vol. 190, Issues 5-6, pp. 363-369, Aug 1994.

[30] W. Wang, T. Ottosson, M. Sternad, A. Ahlén, A. Svensson, "Impact of Multiuser Diversity and Channel Variability on Adaptive OFDM", Proc. IEEE VTC Fall 2003, Oct. 2003.

[31] Mathias Johansson, *Resource Allocation under Uncertainty – Applications in Mobile Communications*, Ph.D. Thesis, Uppsala University, Signals and Systems Group, Oct 2004.
[32] Nilo Casimiro Ericsson, *Revenue Maximization in Resource Allocation – Applications in Wireless Communication Networks*, Ph.D. Thesis, Uppsala University, Signals and Systems Group, Oct 2004.
[33] C. E. Shannon, "A mathematical theory of communication", The Bell System Technical Journal, Vol. 27, pp. 379-423, 623-656, July, October, 1948.
[34] Lewis H. Roberts, "A discipline for the avoidance of unnecessary assumptions", ASTIN Bulletin, Vol. 5, No. 3, Feb 1971.
[35] Sally Floyd, Vern Paxson "Difficulties in simulating the Internet", IEEE/ACM Transactions on Networking, vol. 9, issue 4, pp. 392-403, Aug 2001.
[36] Torbjörn Ekman, *Prediction of Mobile Radio Channels – Modeling and Design*, Ph.D. Thesis, Signals and Systems, Uppsala University, Oct. 2002.
[37] Torbjörn Ekman, Mikael Sternad, Anders Ahlén, "Unbiased power prediction on broadband channels", Proc. IEEE VTC Fall 2002, Sept. 2002.
[38] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view", IEEE Trans. on Communications, Vol. 49, No. 9, pp. 1561-1571, Sept 2001.
[39] J. G. Proakis, *Digital Communications*, third edition, McGraw-Hill, 1995.
[40] Mathias Johansson, "Approximate Bayesian inference by adaptive quantization of the hypothesis space" 25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Aug 7-12 2005.
[41] Anand Bedekar, Sem Borst, Kavita Ramanan, Phil Whiting, Edmund Yeh, "Downlink scheduling in CDMA data networks", Proc. IEEE Globecom'99, Dec 5-9 1999.
[42] S. M. Alamouti, "A simple transmitter diversity Scheme for wireless communications", IEEE J. Selected Areas in Communications, vol. 16, pp. 1451-1458, Oct. 1998.
[43] David N.C. Tse, "Multiuser diversity in wireless networks", Presentation at Stanford University, April 16, 2001.
[44] Mathias Johansson, " Diversity-Enhanced Equal Access – Considerable throughput gains with 1-bit feedback", Proc. IEEE SPAWC 2004, July 2004.
[45] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, fourth edition, Academic Press, 1980.



Mathias Johansson was born in Skövde, Sweden, in 1976. He received the Ph.D. degree in signal processing from Uppsala University, Sweden, in 2004. One of the founders of Dirac Research AB, he is currently working at Dirac Research and Uppsala University. His research interests include Bayesian information processing and decision theory, resource allocation, and audio signal processing.



Mikael Sternad (S'83–M'88–SM'90) Mikael Sternad is Professor in Automatic Control at Uppsala University, Sweden. He received an M.S. degree in engineering physics in 1981 and a PhD in automatic control in 1987, both from the Institute of Technology at Uppsala University, Sweden, and was promoted to professor in 2001. His main research interest is signal processing applied to mobile radio communication problems, such as long-range channel prediction, used for fast link adaptation and scheduling of packet data flows in wireless mobile systems. He is at present involved in the over-all system design of a flexible radio interface for 4G systems within the European Union Integrated project WINNER. His research interests also include acoustic signal processing, in particular compensation of loudspeaker dynamics.