

Analytic Approximation of the Effective Bandwidth for Best-Effort Services in UMTS Networks

Dirk Staehle, Kenji Leibnitz, Klaus Heck, and Phuoc Tran-Gia
Department of Distributed Systems
Institute of Computer Science, University of Würzburg,
Am Hubland, 97074 Würzburg, Germany
[staehle, heck, leibnitz, trangia]@informatik.uni-wuerzburg.de

Bernd Schröder and Albert Weller
T-Mobile International
Karl Duwe-Straße 31, 53227 Bonn, Germany
[Bernd.Schroeder, Albert.Weller]@t-mobile.de

Abstract—The introduction of third generation mobile communication systems allows the service providers to offer a large variety of services which are in the Universal Mobile Telecommunication System subsumed under the categories conversational, streaming, interactive, and background class. While the conversational and streaming classes have a guaranteed bandwidth and delay, the interactive and background class consume the remaining system capacity. On the downlink this system capacity is limited by the base station transmit power and on the uplink by the interference. In our analysis we approximate the capacity which the services with quality of service leave for best-effort services belonging to the interactive class. The background class is neglected. We focus our analysis on the downlink since for UMTS the expected traffic is assumed to be asymmetric with the bulk of it towards the mobile station and derive the distribution of the bandwidth available for best-effort users according to a spatial user distribution.

I. INTRODUCTION

The introduction of third generation mobile communication systems allows the service providers to offer a large variety of services which are in the Universal Mobile Telecommunication System (UMTS) subsumed under the categories conversational, streaming, interactive, and background class, see e.g. [1]. While the conversational and streaming classes have a guaranteed bandwidth and delay, the interactive and background class consume the remaining system capacity. On the downlink this system capacity is limited by the base station (BS) transmit power and on the uplink by the interference. In our analysis we approximate the capacity which the services with quality of service (QoS) leave for best-effort services belonging to the interactive class. The background class is neglected. We focus our analysis on the downlink since the traffic in UMTS networks is expected to be asymmetric with the bulk of it towards the mobile station (MS). On the downlink every BS tries to transmit with a target power which is consumed primarily by common channels with constant power and by users with QoS. The best-effort users share the remaining power. The principle of rate control or rate adaptation by applying dynamic spreading and/or dynamic code allocation with multi-code CDMA is well studied in the literature. In [2] three ways to assign this remaining transmit power to the best-effort users are proposed: equal rates, equal power, and Min/Max rate allocation. Equal rate allocation means that all best-effort users obtain the same rate independent of their

location. This strategy causes that users at the edge of a cell consume considerably more power than users near the BS. In contrast, allocating an equal power to each user causes best-effort users at the edge to have a rather bad performance with little data rates. The Min/Max allocation scheme was originally published in [3]. It serves a single user at a given time instant and optimizes the mean data rates of all best-effort users while maintaining a certain ratio between minimum and maximum rate. In [4], the spreading gain is adapted according to thresholds for the total base station transmit power and the system performance is demonstrated by a simulation. The authors in [5] propose a scheduling scheme for allocating data rates according to the base station power and the channel conditions while guaranteeing a certain quality to all users. In [6] several ways to perform rate control are described. Similar to [2], the paper proposes a scheme with equal rate allocation and a scheme maximizing the total data rate which corresponds to the Min/Max scheme.

In this paper we consider the equal rate allocation strategy since it follows the same principle of fairness used for the QoS users, i.e. all users experience the same quality independent of their radio conditions and their position in the cell. The objective of our work is to derive an analytic method to approximate the quality in terms of bandwidth that a best-effort user experiences in an environment with heterogeneous traffic. The model considers a network of base stations which serve QoS users with different service types and best-effort users. A set of predefined services is available for the best-effort users and their actual service depends on the BS load. In Sec. II we consider the best-effort data rates in a scenario with a given number of users with deterministic positions. In Sec. III the model is extended to stochastic user locations and in Sec. IV we consider stochastic user numbers. The analytic approximations are compared with snapshot simulations to validate their accuracy. Sec. V contains some exemplary numerical results and Sec. VI concludes the paper.

II. DETERMINISTIC USER NUMBER AND POSITION

Consider a set of L BSs and a set of K stationary MSs which either belong to the set Q of MSs with QoS, for which a bit rate R_k and a target $E_b/N_0 \hat{\epsilon}_k^*$ is defined, or to the set B of best-effort MSs. These MSs receive their bit rate with

corresponding target E_b/N_0 values out of a set of possible services depending on the current BS load. The total transmit power \hat{T}_x of BS x comprises a constant part $\hat{T}_{x,C}$ spent for common control channels, a second part $\hat{T}_{x,Q}$ which is dedicated to the QoS MSs, and the remaining part $\hat{T}_{x,B}$ is assigned to the best-effort MSs. The rate control of the UMTS system aims at giving all best-effort MSs of a BS an equal bit rate such that the target BS power $\hat{T}_{x,D}$ is met. In the ideal case, the bit rates are adjusted such that all BSs transmit with equal power $\hat{T}_x = \hat{T}_{x,D}$ which we use in our analysis to approximate the interference from other BSs.

Before starting with the algorithm to compute the distribution of the bit rates some notations are introduced. The variable $\hat{d}_{x,k}$ denotes the signal attenuation from BS x to MS k , the total bandwidth is W , \hat{N}_0 is the thermal noise density, and α is the orthogonality factor. The notation \hat{a} stands for a linear value while a is the corresponding value in decibels. Based on the power control equation

$$\hat{\varepsilon}_k^* = \frac{\frac{W}{R_k} \hat{S}_{x,k} \hat{d}_{x,k}}{W \hat{N}_0 + \left(\sum_{y \neq x} \hat{T}_y \hat{d}_{y,k} + \alpha (\hat{T}_x - \hat{S}_{x,k}) \hat{d}_{x,k} \right)} \quad (1)$$

the transmit power $\hat{S}_{x,k}$ of BS x dedicated to a MS k follows as

$$\hat{S}_{x,k} = \frac{R_k \hat{\varepsilon}_k^*}{W + \alpha R_k \hat{\varepsilon}_k^*} \left(\frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} T_y \frac{\hat{d}_{y,k}}{\hat{d}_{x,y}} + \alpha \hat{T}_x \right) \quad (2)$$

For each service s , with or without QoS, we introduce the service requirement $\omega_s = (R_s \hat{\varepsilon}_s^*) / (W + \alpha R_s \hat{\varepsilon}_s^*)$. As we assume perfect power control, all QoS users of one service have the same service requirement and it remains constant. The best-effort users adapt their service requirement to one of the provided best-effort services according to the system load. At the same instant all best-effort users receive the same quality and hence have the same service requirements. Therefore, we introduce $\omega_{x,B}$ as the service requirement which is common for all best-effort users at BS x . In the following we determine the maximum possible service requirement for best-effort users while maintaining the QoS users' service requirements. The total power of BS x is

$$\hat{T}_x = \hat{T}_{x,C} + \sum_k \omega_k \left(\frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} \hat{T}_y \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} + \alpha \hat{T}_x \right). \quad (3)$$

Under the assumption of perfect power and rate control all BSs z transmit with their target power $\hat{T}_{z,D}$ and we obtain

$$\hat{T}_{x,D} = \hat{T}_{x,C} + \sum_k \omega_k \left(\frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} \hat{T}_{y,D} \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} + \alpha \hat{T}_{x,D} \right). \quad (4)$$

The target BS transmit powers, the service requirements, and the attenuations of the QoS users are given and the transmit

power $\hat{T}_{x,Q}$ required for all QoS users is

$$\hat{T}_{x,Q} = \sum_{k \in Q} \omega_k \left(\frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} \hat{T}_{y,D} \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} + \alpha \hat{T}_{x,D} \right). \quad (5)$$

The transmit power remaining for the best-effort users is then the target transmit power minus the constant power and the power for QoS users

$$\hat{T}_{x,B} = \hat{T}_{x,D} - \hat{T}_{x,C} - \hat{T}_{x,Q} \quad (6)$$

For best-effort users the attenuations are known and the service requirements are equal. Consequently, the service requirements for best-effort users are

$$\omega_{x,B} = \frac{\hat{T}_{x,B}}{\sum_{k \in B} \left(\frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} \hat{T}_{y,D} \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} + \alpha \hat{T}_{x,D} \right)}. \quad (7)$$

For reasons of readability we introduce the following notations:

$$Z_k = \frac{W \hat{N}_0}{\hat{d}_{x,k}} + \sum_{y \neq x} \hat{T}_{y,D} \frac{\hat{d}_{y,k}}{\hat{d}_{x,k}} \quad (8)$$

$$H_{x,B} = \sum_{k \in B} \left(Z_k + \alpha \hat{T}_{x,D} \right). \quad (9)$$

The obtained service requirements are optimal and applying them demands that the available service requirements are continuous. Actually, only a predefined set of services is available. A service s has a predefined bit rate $R_{s,B}$ and target- E_b/N_0 value $\varepsilon_{s,B}^*$ and the adoptable service requirements are discrete. Therefore, the rate control assigns the largest rate such that the best-effort service requirements remain below the optimal service requirements. We consider two sets of offered best-effort services with a basic bit rate of 16kbps. The first set uses dynamic spreading, i.e. the bit rates are doubled or decreased by half. The maximum rate is 256kbps such that 5 services with 16kbps, 32kbps, 64kbps, 128kbps and 256kbps are available. The second set additionally allows multiple codes per user such that all multiples of 16kbps up to 256kbps are possible. The target- E_b/N_0 is 3dB for all services. Optimal best-effort rates R_{opt} are computed from the optimal service requirements $\omega_{x,B}$ by

$$R_{opt} = \begin{cases} R_{1,B}, & \text{if } \omega_{x,B} < \omega_{1,B} \\ \frac{\omega_{x,B} W}{\varepsilon_B^* - W \alpha \varepsilon_B^*}, & \text{if } \omega_{1,B} \leq \omega_{x,B} < \omega_{max,B} \\ R_{max,B}, & \text{if } \omega_{x,B} \geq \omega_{max,B} \end{cases}, \quad (10)$$

whit max denoting the maximum service. The direct mapping of optimal service requirements to optimal bit rates is possible only if the target- E_b/N_0 is constant and independent of the bit rate. In Fig. 1 the bit rates corresponding to the computed optimal service requirements are depicted.

The optimal service requirements for the deterministic model with given mobile number and position are computed according to Eqn. (7) and the resulting bit rates are obtained according to Fig. 1. We denote a set of mobiles with positions that are generated according to a spatial point process as a snapshot. We use the distributions resulting from a series of

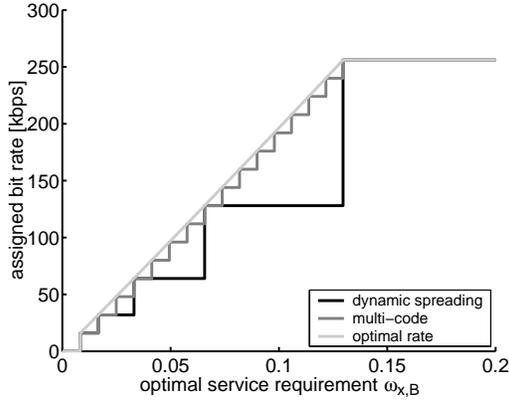


Fig. 1. Assignment of bit rates to optimal service requirements

snapshots for validating the analytical approximation of the optimal service requirements and corresponding bit rates.

III. STOCHASTIC USER POSITION

In this section we consider a deterministic number of users with stochastic positions and, consequently, stochastic attenuations. The variable $\bar{n} = (n_1, \dots, n_S, n_B)$ denotes the number of users where n_s is the number of users with service s and S services with guaranteed quality are considered. The number of best-effort users in the system is n_B . In the following we assume that for a given user number \bar{n} the optimal service requirement $\omega_{x,B}(\bar{n})$ follows a lognormal distribution and derive the first and second moment.

Taking the logarithm of Eqn. (7) yields

$$\log(\omega_{x,B}(\bar{n})) = \log(\hat{T}_{x,B}(\bar{n})) - \log(H_{x,B}(\bar{n})). \quad (11)$$

If $\omega_{x,B}(\bar{n})$ is a lognormal random variable the logarithm of $\omega_{x,B}(\bar{n})$ follows a normal distribution with

$$E[\log(\omega_{x,B}(\bar{n}))] = E[\log(\hat{T}_{x,B}(\bar{n}))] - E[\log(H_{x,B}(\bar{n}))], \quad (12)$$

$$\text{VAR}[\log(\omega_{x,B}(\bar{n}))] = \text{VAR}[\log(\hat{T}_{x,B}(\bar{n}))] + \text{VAR}[\log(H_{x,B}(\bar{n}))]. \quad (13)$$

The mean and variance of $\hat{T}_{x,B}(\bar{n})$ are:

$$E[\hat{T}_{x,B}(\bar{n})] = \hat{T}_{x,D} - \hat{T}_{x,C} - \sum_{s=1}^S n_s \omega_s (E[Z] + \alpha \hat{T}_{x,D}) \quad (14)$$

$$\text{VAR}[\hat{T}_{x,B}(\bar{n})] = \sum_{s=1}^S n_s \omega_s^2 \text{VAR}[Z], \quad (15)$$

where Z is a iid random variable for Z_k which depends solely on the location of a user in the cell. The first and second moment of Z are determined by numerical integration over the plane. Let F be a set of points in the considered cell then

$$E[Z^t] = \sum_{f \in F} p(f) \left(\frac{W \hat{N}_0}{\hat{d}_{x,f}} + \sum_{y \neq x} \hat{T}_{y,D} \frac{\hat{d}_{y,f}}{\hat{d}_{x,f}} \right)^t, \quad (16)$$

with $p(f)$ the traffic density at point f and $\sum_{f \in F} p(f) = 1$. The mean and variance of $H_{x,B}$ also depend on the moments of Z and are:

$$E[H_{x,B}(\bar{n})] = n_B (E[Z] + \alpha \hat{T}_{x,D}) \quad (17)$$

$$\text{VAR}[H_{x,B}(\bar{n})] = n_B \text{VAR}[Z] \quad (18)$$

Assuming that both $\hat{T}_{x,B}$ and $H_{x,B}$ are approximately lognormal, the logarithms of these random variables both follow a normal distribution. Generally, for a lognormal random variable X the mean and the variance of $\log(X)$ are

$$\begin{aligned} E[\log(X)] &= 2 \log(E[X]) - \frac{1}{2} \log(E[X^2]) \\ \text{VAR}[\log(X)] &= \log(E[X^2]) - 2 \log(E[X]). \end{aligned} \quad (19)$$

Accordingly, the mean and the variance of the logarithms of $\hat{T}_{x,B}$ and $H_{x,B}$ are computed, and Eqn. (12) and Eqn. (13) yield the moments of $\log(\omega_{x,B}(\bar{n}))$. The moments of $\omega_{x,B}(\bar{n})$ follow by the inverse computation of Eqn. (19):

$$E[X^t] = \exp(tE[\log(X)] + \frac{1}{2}t^2 \text{VAR}[\log(X)]) \quad (20)$$

Fig. 2 shows the CDFs of the optimal service requirements $\omega_{x,B}(\bar{n})$ for different user configurations. The solid lines correspond to the CDF obtained by generating a series of snapshots and the dashed lines depict the CDFs of the approximated lognormal distributions. The example considers n_Q QoS users with a bit rate of 64kbps and a target- E_b/N_0 of 4 dB and the best-effort users can adopt the services of the set with dynamic spreading and without multi-code. The users are homogeneously distributed in a hexagonal area around a central BS and another 38 BS are arranged in a hexagonal grid around the central BS with a distance of 2km between each two BSs. The pathloss model uses the formula given in [7] with $d_{x,k} = -128.1 - 37.6 \log_{10}(\text{dist}_{x,k})$ where $\text{dist}_{x,k}$ is the distance from BS x to MS k in km. The target BS transmit power is set to 10W for all BSs and the constant part of the transmit power is 2W. The CDFs correspond to different mixtures of QoS and best-effort users. The lognormal approximations match well with the snapshot simulation. The

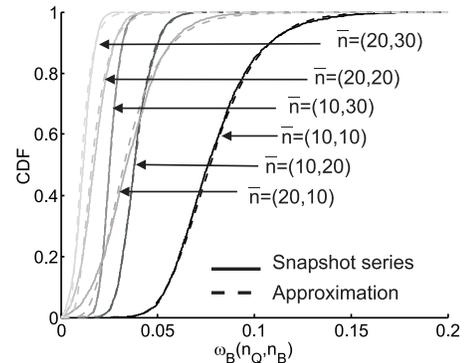


Fig. 2. Approximated and Snapshot CDFs of optimal service requirements for various user configurations with dynamic spreading

probabilities for very small optimal service requirements are partially underestimated by the approximation. However, the error is negligible since the CDFs match well for values above the requirement for the smallest service. It is worth noting that the number of QoS users dominates the mean of the optimal service requirements while the number of best-effort users dominates the variance.

Fig. 2 shows the distribution of the optimal service requirements and Fig. 1 allows to map the distribution to the bit rate or service probability of the best-effort users. The probability $p(s, \bar{n})$ that with \bar{n} users in the system the best-effort users can operate with service s is

$$p(s|\bar{n}) = \begin{cases} P\{\omega_{x,B}(\bar{n}) < \omega_{s,B}\} & \text{for } s = 0 \\ P\{\omega_{s,B} \leq \omega_{x,B}(\bar{n}) < \omega_{s+1,B}\} & \text{for } 1 \leq s < S \\ P\{\omega_{S,B} \leq \omega_{x,B}(\bar{n})\} & \text{for } s = S \end{cases} \quad (21)$$

where $s = 0$ means that no service is possible. Fig. 3 illustrates how to derive the service probabilities at the example of two user configurations from Fig. 2. The vertical lines show the values of ω_s for the available best-effort services. The horizontal lines indicate the probability that a service is not possible for the user configuration. Thus, the intervals between the vertical lines correspond to the service probabilities. For example, with 20 QoS users and 20 best-effort users with about 6% no service is available. The probability for service 1 corresponds to the interval to the next higher line and is about 47.5%. Services 4 and 5 are impossible in this state. On the other hand, for the user configuration with lower load, service 5 is taken with 2.5%, service 4 dominates with over 70% and the QoS users always leave enough capacity for service 3.

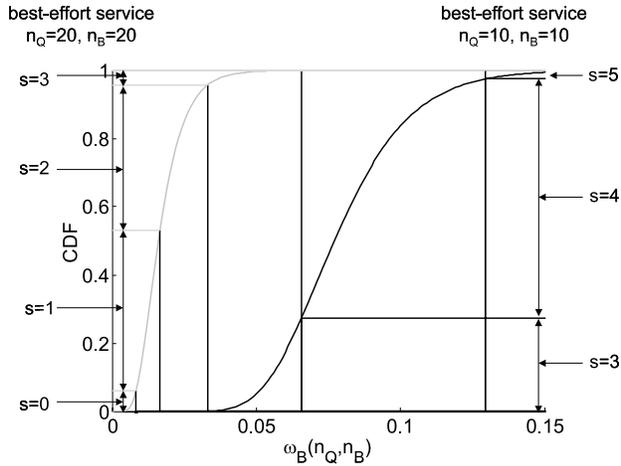


Fig. 3. Determination of the service probabilities from the CDF of $\omega_B(\bar{n})$

IV. STOCHASTIC USER NUMBER PER SERVICE

Eqn. (21) yields the service distribution for a deterministic user configuration \bar{n} . The objective of this analysis, however, is to obtain the best-effort service probabilities for a spatial traffic distribution with a certain user density. The average

number of QoS users with service s in the cell area is a_s and the mean number of best-effort users is a_B . The vector $a = (a_1, \dots, a_S, a_B)$ denotes the total traffic load. We assume that both the number of QoS and best-effort users follow a Poisson distribution. The probability of a deterministic user configuration \bar{n} follows from a modified product form solution with

$$\tilde{p}(\bar{n}) = \phi(\bar{n}) \prod_{s=1}^S \frac{a_s^{n_s} a_B^{n_B}}{n_s! n_B!} \quad \text{and} \quad (22)$$

$$p(\bar{n}) = \frac{\tilde{p}(\bar{n})}{\sum_{\bar{n}' \in \Omega} \tilde{p}(\bar{n}')}. \quad (23)$$

The variable Ω denotes the set of all user configurations and is defined as

$$\Omega = \left\{ \bar{n} \mid \alpha \left(\sum_{s=1}^S n_s \omega_s + n_B \omega_{1,B} \right) < 1 \right\}. \quad (24)$$

In the analysis a user configuration is considered only if the target transmit power is sufficient to provide the service requirement of the QoS users. For all user configurations $\bar{n} \in \Omega$ the target transmit power can be sufficient to support all QoS users and the best-effort users with service 1. However, there are also spatial user locations such that the transmit power is not sufficient for the smallest best-effort service and they occur with probability $p(0, \bar{n})$. Furthermore, the target transmit power may even be insufficient to support all QoS users. The probability of such a spatial user pattern is $1 - \phi(\bar{n})$ and accordingly $\phi(\bar{n})$ is the probability that the spatial distribution of the QoS users allows a solution of the power control equation and is given by

$$\phi(\bar{n}) = P \left\{ \left(1 - \alpha \sum_{s=1}^S n_s \right) \hat{T}_{x,D} - \hat{T}_{x,C} < H_{x,Q}(\bar{n}) \right\}, \quad (25)$$

with $H_{x,Q}(\bar{n}) = \sum_{k \in Q} \omega_k Z_k$. Assuming $H_{x,Q}(\bar{n})$ as a lognormal random variable, $\phi(\bar{n})$ directly follows from the CDF. The moments of $H_{x,Q}(\bar{n})$ are evident.

The probability $p(s)$ that according to an underlying spatial traffic distribution the best-effort users operate with service s results from the theorem of total probability:

$$p(s) = \frac{\sum_{\bar{n} \in \Omega'} p(\bar{n}) p(s|\bar{n})}{\sum_{\bar{n} \in \Omega'} p(\bar{n})}, \quad (26)$$

with $\Omega' = \{\bar{n} \in \Omega \mid n_B > 0\}$. In Fig. 4 the service distributions for different user densities are shown. As in the previous example the best-effort services with dynamic spreading and without multi-code are considered. The user densities reach from a low load with an average of five 64kbps QoS users and ten best-effort users to a high load with 20 QoS users and 30 best-effort users in mean. The brightness of the curves corresponds to the traffic system load with lighter colors for lower load. Again, the approximations are validated by snapshot simulation and the results match well for all traffic densities. For the lowest load, service 5 with 256kbps is available to about 28% and the probability for exceeding the

requirements of service 2 is negligible. The opposite occurs for the high load. Services 4 and 5 are never adopted and with up to 30% the best-effort users are not served at all. For medium loads with 10 to 20 QoS users and 10 to 20 best-effort users mainly services 1, 2 and 3 are taken.

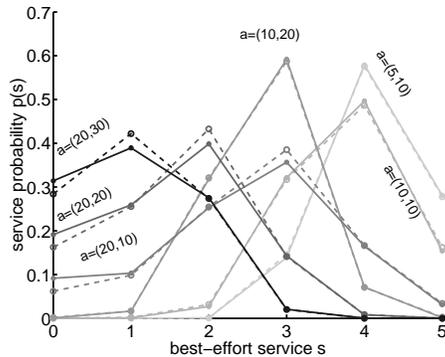


Fig. 4. Probabilities of best-effort services

V. NUMERICAL RESULTS

In this section we compare the performance of dynamic spreading and multi-code usage with the optimal best-effort rate allocation. Fig. 5 shows the cumulative distribution of the best-effort bit rates R_s for a low traffic load with $a = (10, 10)$ and a high load with $a = (20, 20)$. The curves for all three best-effort service sets agree with each other for the bit rates of dynamic spreading. Obviously, the optimal rate assignment yields the highest data rate but using multiple codes leads to a near optimum performance in particular for the low load scenario. Dynamic spreading alone shows a considerably worse performance.

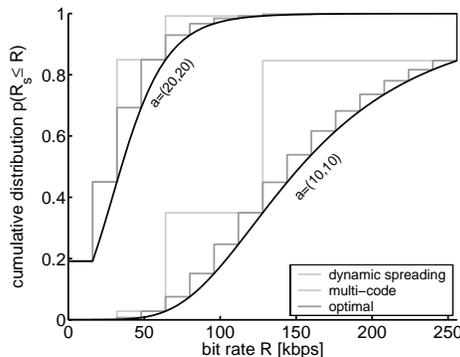


Fig. 5. Cumulative distribution of best-effort data rates with dynamic spreading and multi-code

Fig. 6 confirms this behavior. It shows the average data rates for different mean numbers of QoS and best-effort users. For high loads, the curves are almost identical while for low

loads the usage of multiple codes increases the average best-effort data rate by up to 35kbps. The relative gain, i.e. the ratio between mean data rates with and without multi-code, reaches from over 20% for 10 best-effort users or less than 10 QoS users in mean to 5% for an average of 30 best-effort and 25 QoS users.

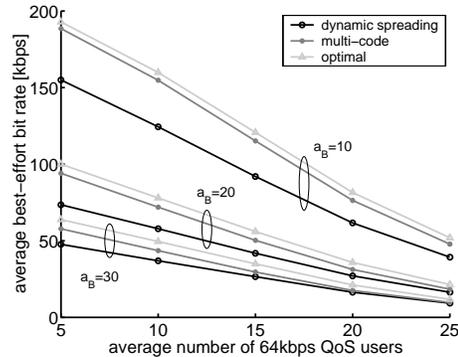


Fig. 6. Average best-effort data rates for different traffic mixtures

VI. CONCLUSION

We have presented an analytic model to determine probabilities for the quality provided to a best-effort user. The approximation is based on the assumption that due to the rate control all BS transmit with a predefined and constant power. In the planning process of 3G networks the approximation of the effective bit rates is necessary to provide a certain, though not guaranteed, quality for best-effort users. The results from our model match very well with the results from the snapshot simulation. We have shown that the usage of multiple codes improves the performance of the rate control considerably compared to dynamic spreading.

REFERENCES

- [1] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimisation for UMTS*. John Wiley & Sons, Ltd., 2002.
- [2] H. Kröner, "Radio resource allocation for data services in umts networks," *Internation Journal of Electronics and Communications (AEÜ)*, vol. 55, no. 1, pp. 1–8, 2001.
- [3] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Comm. Mag.*, vol. 38, no. 7, pp. 70–77, Jul 2000.
- [4] W. Rave, T. Köhler, G. Fettweis, P. Schneider, and M. Berg, "Evaluation of load control strategies in an UTRA/FDD network," in *Proc. of VTC Spring*, Rhodes, Greece, May 2001, pp. 2710–2714.
- [5] L. Xu, X. Shen, and J. Mark, "Dynamic bandwidth allocation with fair scheduling for WCDMA systems," *IEEE Wireless Comm.*, vol. 9, no. 2, pp. 26–32, Apr 2002.
- [6] C. Mihailescu, X. Lagrange, and P. Godlewski, "Radio resource management for packet transmission in UMTS WCDMA systems," in *Proc. of VTC Fall'99*, Amsterdam, The Netherlands, Sep 1999, pp. 573–577.
- [7] ETSI, "Selection procedures for the choice of radio transmission technologies of the UMTS, Tech. Rep. TR 101 112 V3.2.0, 1998.