

Service Differentiation in Third Generation Mobile Networks

Vasilios A. Siris^{a,*}, Bob Briscoe^b, and Dave Songhurst^b

^a Institute of Computer Science, FORTH, Crete, Greece

^b BT Research, Adastral Park, Ipswich, UK

12 May, 2002

Abstract

We present and analyse an approach to service differentiation in third generation mobile networks based on Wideband CDMA, that exposes a new weighting parameter designed to reflect allocation of the congestible resource. The approach naturally takes into account the difference in resource scarcity for the uplink and downlink, because it is grounded on fundamental economic models for efficient utilization of resources in WCDMA. If required, discrete values of this weight parameter can be presented as different service classes. Finally we present numerical experiments demonstrating the effectiveness of our approach, and investigate its performance and transient behaviour under power control and signal quality estimation errors.

Keywords: resource usage, class-based, power control, congestion control

1 Introduction

The percentage of users accessing packet switched networks through wireless access networks is increasing at a very large pace. Hence, the ability to support quality of service (QoS) differentiation in wireless systems is becoming increasingly important. Indeed, the UMTS (Universal Mobile Telecommunication System) third generation mobile telecommunication system allows user negotiation of bearer service characteristics, such as throughput, error rate, delay, etc [1, 4]. WCDMA (Wideband Code Division Multiple Access) is the main air interface for UMTS. WCDMA is based on Direct Sequence CDMA (DS-CDMA), a spread spectrum technology where data bits are spread over the entire spectrum used for transmission. In DS-CDMA all mobile users can simultaneously transmit utilizing the whole radio spectrum, and unique digital codes are used to differentiate the signal from different mobiles. Variable bit rates are achieved using variable spreading factors, where the spreading factor determines how much a data bit is spread in time, and multiple codes. The signal quality (error rate) is determined by the ratio of the power of the received signal over the total interference. Because all mobile users can simultaneously transmit data, the interference a mobile encounters includes, in addition to noise, interference due to signals from other mobiles.

In this paper we propose models and procedures for service differentiation in WCDMA networks. By considering actual resource usage in both the uplink and the downlink, our procedures are fair and efficient, and are robust to varying demand for wireless resources. In the uplink resource usage is radio spectrum limited, being an increasing function of the product of the transmission rate and the signal quality, the latter expressed in terms of the signal-to-interference ratio, SIR. In the downlink resource usage is constrained by the total transmission power at the base

*Corresponding author: Institute of Computer Science, FORTH, P.O. Box 1385, Heraklion GR 71 110, Crete, Greece. Tel.: +30 810 391726, email: vsiris@ics.forth.gr

station. Among our objectives is to discuss how the proposed procedures relate and interoperate with resource management functions already defined for WCDMA. In particular, our approach involves changes to the procedures running on the RNC, and in particular procedures for outer loop power control and load control, while they do not change fast closed-loop power control, which operates on a much faster timescale and is implemented in the physical layer.

Regarding related work, the authors of [3] present a class-based quality of service framework. The performance for different classes depends on an elasticity associated with each class; this elasticity indicates how the rate of each class will decrease in periods of congestion. In the downlink, in addition to the elasticity, the path loss is also taken into account for allocating resources. Our approach differs from the above in that allocation of resources is done proportional to weights, thus leading to fair, in terms of weights, allocations. Moreover, by charging in proportion to weights the approach can lead to fair charging. The authors of [5] provide an overview radio resource allocation techniques, focusing on power and rate adaptation in overload periods. The work in [6] discusses rate adaptation for different wireless technologies.

The rest of the paper is organized as follows. In Section 2 we discuss resource management procedures in WCDMA. In Section 3 we first discuss resource usage in the uplink and downlink, and then propose procedures for service differentiation in each direction. In Section 4 we present and discuss numerical investigations demonstrating the effectiveness of our approach, and in Section 5 we conclude the paper identifying related and future research issues.

2 Resource management in WCDMA

Among the main resource management functions in WCDMA are the following [4]:

- fast closed-loop power control
- outer loop power control
- load control

In the uplink, with fast closed-loop power control, Figure 1(a), the base station continuously measures the received SIR for each user, and then compares the measured SIR with a target SIR . If the measured SIR is smaller, then the base station instructs the mobile to increase the transmission power. On the other hand, if the measured SIR is smaller than the target SIR , then the base station instructs the mobile to decrease its transmission power. The increase and decrease commands are binary, i.e., each power update results in an increase or decrease of the transmission power by a fixed amount, typically in the range 0.5 – 1 dB. The above closed loop

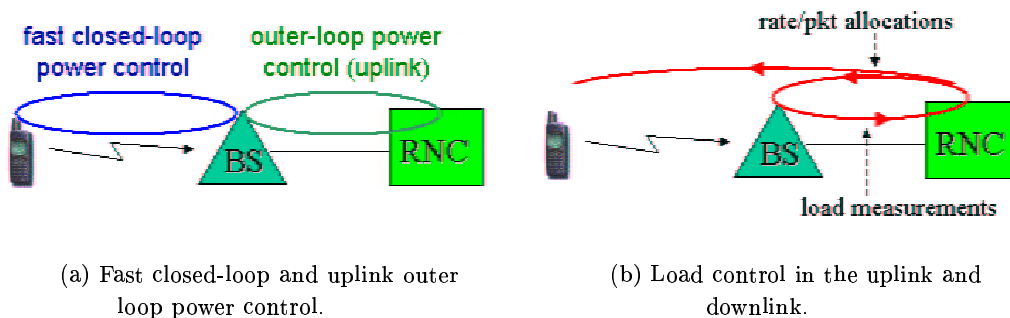


Figure 1: Resource management functions in WCDMA.

control cycle operates at a frequency of 1500 Hz, which corresponds to one power update every 0.67 msec. In WCDMA, a similar fast closed-loop power control loop exists in the downlink direction, where now the *SIR* is measured by the mobile, and power update commands are sent by the mobile to the base station.

In second generation systems, the main objective of fast closed-loop power control is to tackle the *near-far* problem: In the case of two mobile hosts, one very close and one very far from the base station, if both mobiles had the same transmission power then the signal from the mobile nearer to the base station, due to the smaller path loss, would overwhelm the signal from the more distant mobile. Fast closed-loop power control can resolve this problem, by maintaining the same *SIR* for all mobiles. In third generation system, which will support applications with different quality of service requirements, the target *SIR* need not be the same, since a different target *SIR* yields a different signal quality, in terms of the frame error rate, *FER*. Typically, for non-real-time services the frame error rate is in the range 10% – 20%, whereas for real-time services it is close to 1% [4, p. 193].

Due to differences in the propagation environment and the mobile speed, there is no one-to-one correspondence between target *SIR* and the achieved *FER*. For this reason there is a need for outer loop power control. In the uplink, outer loop power control operates between the base station and the radio network controller (RNC), Figure 1(a), and its objective is to adjust the target *SIR* in order to achieve some predefined frame error rate. In the downlink, outer loop power control operates within the mobile host. Typically, outer loop power control operates in timescales slower than those of fast closed-loop power control.

In WCDMA data transmission occurs in fixed-size frames, which have a minimum duration of 10 msec; the rate is allowed to change between frames but remains constant within a single frame. Hence, rate control occurs in timescale much slower than the timescales of fast closed-loop power control, where one power update occurs every 0.67 msec. Moreover, WCDMA supports discrete bit rates: in the uplink the user data rate can obtain the values 7.5, 15, 30, 60, 120, 240, 480 Kbps, which correspond to a spreading factor of 256, 128, 64, 32, 8, 4, respectively [4]. Higher bit rates can be achieved with the use of multiple codes. Finally, in addition to code-division scheduling, WCDMA supports time-division scheduling, that controls which mobiles can transmit in each frame.

The last control procedure we consider is load control, whose objective is to reduce the load during periods of overload. Such periods are detected by the RNC based on measurements it receives from the base station, Figure 1(b). Load control actions include decreasing the target signal-to-interference ratio, *SIR*, that is used in fast closed-loop power control, or decreasing the transmission rate and/or adjust the time scheduling of data, for both the uplink and the downlink.

3 Models for service differentiation

In this section we first discuss resource usage in the uplink and downlink of CDMA systems. Then we present our approach for service differentiation, which involves allocating resources to individual mobile users based on weights. Different weights can be associated with different classes of service offered by the wireless network operator. An important property of CDMA networks is that there are two control parameters that affect a mobile user's level of service: the transmission rate and the signal-to-interference ratio. However, resource usage in the two directions is different, leading to different models for service differentiation.

3.1 Resource usage in CDMA

3.1.1 Resource usage in the uplink

In the uplink, the signal-to-interference ratio at the base station for the transmission from mobile i is given by [2, 9]

$$SIR_i = \frac{W}{r_i} \frac{g_i p_i}{\sum_{j \neq i} g_j p_j + \eta}, \quad (1)$$

where W is the chip rate, which is fixed and equal to 3.84 Mcps for WCDMA, r_i is the transmission rate, p_i is the transmission power, g_i is the path gain between the base station and mobile i , and η is the power of the background noise at the base station. The ratio W/r_i is the spreading factor or processing gain for mobile i .

The value SIR corresponds to the signal quality, since it determines the bit error rate, BER [2, 9], and hence the frame error rate, FER . If we assume perfect power control, then the value of SIR in (1) will be equal to the target signal-to-interference ratio, SIR^* , that is used in fast closed-loop power control.

Solving the set of equations given by (1) for each mobile i , we get [9, 7]

$$g_i p_i = \frac{\eta \alpha_i^{UL}}{1 - \sum_j \alpha_j^{UL}}, \quad \text{where} \quad \alpha_i^{UL} = \frac{1}{\left(\frac{W}{r_i SIR_i} + 1\right)}. \quad (2)$$

Since the power p_i can take only positive values, from (2) we get

$$\sum_i \alpha_i^{UL} < 1. \quad (3)$$

The last equation illustrates that the uplink is *interference-limited*: Even when they have no power constraints, mobile hosts cannot increase their power with no bound, due to the increased interference they would cause to the other mobiles. If (3) is violated, then the target SIR values cannot be met for all mobiles.

A useful expression for measuring the uplink load factor $\sum_i \alpha_i^{UL}$ can be found by summing (2) for all mobiles:

$$\sum_i g_i p_i = \frac{\eta \sum_i \alpha_i^{UL}}{1 - \sum_j \alpha_j^{UL}} \Rightarrow \sum_i \alpha_i^{UL} = \frac{I_{total} - \eta}{I_{total}}, \quad (4)$$

where I_{total} is the total received power, including the noise power. Hence, the estimation of the uplink load factor requires measurements of the total interference and the noise, both of which can be performed at the base station.

In practise, due to limited transmission power from the mobile hosts, imperfect power control, shadowing, etc, the total load must be well below 1. Indeed, in radio network planning [4], all the above factors are used to determine an interference margin (or noise rise) I_{margin} , which is an upper bound on the ratio of the total received power over the noise $\frac{I_{total}}{\eta} \leq I_{margin}$, based on which (4) becomes

$$\sum_i \alpha_i^{UL} \leq \frac{I_{margin} - 1}{I_{margin}}. \quad (5)$$

Next we discuss how the interference margin I_{margin} can be estimated from various parameters used in radio network planning. First assume that all mobiles have the same target SIR^* , spreading factor W/r , and power limit \bar{p} . The mobile that incurs the largest path loss, or equivalently the smallest gain g_{min} , will determine the tightest resource constraint, hence from (1) we have

$$SIR^* \leq \frac{W}{r} \frac{g_{min} \bar{p}}{I_{max}} \Rightarrow I_{max} \leq \frac{W}{r} \frac{g_{min} \bar{p}}{SIR^*},$$

which in dB can be written as

$$I_{max,dBm} = \left(\frac{W}{r}\right)_{dB} + g_{min,dB} + \bar{p}_{dBm} - SIR_{dB}.$$

The last equation can be modified to take into account WCDMA-specific parameters that are used in radio network planning, which include the fast fading margin, *FFM*, the slow fading margin, *SFM*, and the soft handover gain, *SHG* [4]. If we consider these parameters, then the last equation becomes

$$I_{max,dBm} = \left(\frac{W}{r}\right)_{dB} + g_{min,dB} + \bar{p}_{dBm} - SIR_{dB} - FFM_{dB} - SFM_{dB} + SHG_{dB}.$$

If the signal-to-interference ratio *SIR* and the transmission rate *r* are not the same for all mobiles, then we can replace the parameters in the last equation with $SIR_{max,dB}$ and r_{max} . Moreover, a more detailed analysis can also consider factors such as antenna gains and losses internal to the mobile and the base station. Finally, we note that the above results can be generalized for the case of multiple cells by including the intercell coefficient $f = (\text{other cell interference})/(\text{intracell interference})$ [2, 4], in which case (5) becomes

$$\sum_i \alpha_i^{UL} \leq \frac{1}{1+f} \frac{I_{margin} - 1}{I_{margin}}.$$

When there are a large number of mobile users, each using a small portion of the available resources, we have $\frac{W}{r_i SIR_i} \gg 1$, hence $\alpha_i^{UL} \approx \frac{r_i SIR_i}{W}$ and the resource constraint (5) can be approximated by

$$\sum_i r_i SIR_i \leq \rho^{UL} W, \quad \text{where } \rho^{UL} = \frac{I_{margin} - 1}{I_{margin}}. \quad (6)$$

3.1.2 Resource usage in the downlink

In the downlink, the signal-to-interference ratio at mobile *i* is

$$SIR_i = \frac{W}{r_i} \frac{g_i p_i}{\theta_i g_i \sum_{j \neq i} p_j + \eta_i}, \quad (7)$$

where r_i is the transmission rate, p_i is the transmission power, g_i is the path gain between the base station and mobile *i*, θ_i is the orthogonality factor for the codes used in the downlink, and η_i is the power of the background noise at mobile *i*. The orthogonality factor θ_i depends on multipath effects, hence can be different for different mobiles.

In the downlink, the total power with which a base station can transmit has an upper limit, say *P*. Hence, the downlink is *power-limited* and resource usage is determined by the transmission power. On the other hand, recall that the uplink is *interference-limited* and resource usage is determined by the product $r_i SIR_i$. As with the uplink, in the downlink the utilization in practise cannot reach 100%. Hence, the the resource constraint in the downlink is

$$\sum_i p_i \leq \rho^{DL} P, \quad (8)$$

where the downlink utilization factor ρ^{DL} can be estimated in a manner similar to how the uplink utilization factor is estimated.

3.2 Service differentiation in the uplink

In this section we discuss service differentiation in the uplink. Assume that each mobile user has an associated weight; this weight can correspond to a service class selected by the mobile user. To achieve fair resource allocation, wireless resources should be allocated in proportion weights. Due to such proportional allocation, and since resource usage in the uplink is given by the product of the transmission rate and signal-to-interference ratio, from (6) we have

$$r_i SIR_i = \frac{w_i}{\sum_j w_j} \rho^{UL} W. \quad (9)$$

Recall that users can potentially control both the transmission rate and signal-to-interference ratio. How this selection is done depends on what the user values. Indeed, for users that value only the average throughput of data transmission, one can prove that the optimal signal-to-interference ratio depends solely on the frame error rate as a function of the signal-to-interference ratio, and is independent of the transmission rate [8]. In this case, (9) can be used to compute the transmission rate as follows

$$r_i = \frac{1}{SIR_i} \frac{w_i}{\sum_j w_j} \rho^{UL} W. \quad (10)$$

The application of the above equation would be part of load control. Equation (10) can also be applied in the case of traffic that is rate adaptive, but has fixed quality of service requirements, in terms of the frame error rate, *FER*. In this case, outer loop power control would be responsible for adjusting the target *SIR*, and load control would be responsible for adjusting the transmission rate using (10).

The application of (10) allows two alternatives regarding the value for the signal-to-interference ratio that appears in the denominator of the right-hand side: SIR_i can be either the target *SIR* for mobile i , or it can be the actual *SIR* for mobile i , which is estimated from (1). We investigate these two alternatives in Section 4.

In the case of traffic with fixed-rate requirements, but which is adaptive to the signal quality, then based on (9) we can use the following equation for allocating resources (*SIR* values) in proportion to weights:

$$SIR_i = \frac{1}{r_i} \frac{w_i}{\sum_j w_j} \rho W.$$

Application of this equation would affect outer loop power control.

Although the models and procedures discussed above lead to simple (proportional) allocation of resources, they can be theoretically justified in terms of economically efficient resource usage [8]. Indeed, mobile users having a fixed weight correspond to users with a logarithmic utility function; a user's utility function represents how much he values a particular level of service. If each user is charged in proportion to his weight, equivalently in proportion to the amount of resources he is receiving, then the resulting allocations in the equilibrium maximize the aggregate utility of all users (social welfare). If the utility of a user i has a more general form $U_i(x_i)$, where $x_i = r_i(1 - FER(SIR))$ is the average throughput, then the user's weight can be modified slowly according to $w_i = U'_i(x_i)x_i$.

3.3 Service differentiation in the downlink

In the downlink the resource constraint is related to the total transmission power, (8). Based on this equation, if we are to allocate power levels in proportion to weights we have the following

$$p_i = \frac{w_i}{\sum_j w_j} \rho^{DL} P. \quad (11)$$

There are two alternatives for applying the last equation: The first alternative is to have the base station select the instantaneous transmission power based directly on (11). The second alternative is to use (11) for determining an average power, which is then used to compute the transmission rate.

Due to multipath fading, the first alternative has the disadvantage that the received signal quality at a mobile host is not constant. Moreover, it requires modification of the fast closed-loop power control procedure, which is implemented in the physical layer of CDMA systems.

The second alternative can be applied as follows. Let \bar{p}_i be the average power for user i . Based on (11), the average power will be

$$\bar{p}_i = \frac{w_i}{\sum_j w_j} \rho^{\text{DL}} P.$$

As was the case in the uplink, if users value only the average throughput of data transmission, then the optimal signal-to-interference ratio is independent of the transmission rate. Hence, from (7), and assuming the base station achieves the maximum power utilization, the transmission rate for user i will be

$$r_i = \frac{W}{SIR_i} \frac{g_i}{\theta_i g_i \rho^{\text{DL}} P + \eta_i} \frac{w_i}{\sum_j w_j} \rho^{\text{DL}} P. \quad (12)$$

Note that, if mobile i is moving, it is appropriate to take average values for the path gain and the orthogonality factor. From the last equation observe that, as was the case for the power, the transmission rate is proportional to the weight factor.

Equation (12) requires estimation of the channel gain from the base station to the mobile, which can be done using the pilot bits in the downlink control channel. There are two alternatives as to where the selection of r_i based on (12) is performed: the mobile host or the radio network controller (RNC). The first alternative results in more complexity at mobile hosts. Moreover, it requires communicating the parameters $\rho^{\text{DL}} P$ (when the mobile connects to a new base station) and $\sum_j w_j$ (when the sum changes with the arrival or departure of mobile users) from the RNC to the mobile. On the other hand, if the RNC performs the selection, then there would be increased signalling overhead between the mobile and the RNC, since the values of the gain and the orthogonality factor would need to be communicated to the RNC whenever they changed; how often these parameters change depends on the mobile's movement.

It is interesting to observe from (12) that, for the same weight factor, a smaller path gain will result in a smaller (worse) signal quality. Hence, such an allocation differentiates users based on their distance from the base station. Note, however, that when the noise is very small, the distance does not affect the signal quality.

To avoid differentiation due to a mobile's position, one can use (12), after replacing the parameters with their corresponding averages *over all mobile hosts*. Hence, if $\bar{\theta}$ is the average orthogonality, $\bar{\eta}$ is the average noise, and \bar{l} is the average loss, where the loss for user i is $l_i = 1/g_i$, the allocation of rates can use the following equation

$$r_i = \frac{W}{SIR_i} \frac{1}{\bar{\theta} \rho^{\text{DL}} P + \bar{l} \bar{\eta}} \frac{w_i}{\sum_j w_j} \rho^{\text{DL}} P.$$

The decision of whether to allocate resources in the downlink with or without dependence on the mobile's position will be determined by a wireless operator's policy.

4 Numerical investigations

In this section we present simulation investigations that demonstrate the effectiveness of our approach, and investigate how various characteristics of wireless systems,

parameter	values
noise, η	10^{-13} Watt
mobile power, \bar{p}	250 mW
interference margin, I_{margin}	3 dB
intercell interference, f	0
path gain, $g(d) = kd^{-u}$	$u = 3.52, k = 1.82 \cdot 10^{-14}$
target SIR^*	5
power control error, PCE	0 or 1 dB
SIR estimation error, SIR_{err}	0 or 1 dB
# of mobiles, N	11

such as power control and SIR estimation errors, and discrete transmission rates, affect service differentiation and the system's transient behaviour.

The simulation parameters are shown in Table 4. Both the power control and SIR estimation errors are assumed to be lognormally distributed. We assume that the start time for each mobile is randomly distributed in the interval $[0, \lceil N/2 \rceil]$, where N is the total number of mobiles. The simulation experiments were performed in MATLAB, and utilized some functionality of the RUNE library [10].

We have performed experiments for both the uplink and the downlink. Due to space limitations, we present and discuss experiments for the uplink, where rate allocation is based on (10). There are two alternatives for applying the last equation that refer to the value for the signal-to-interference ratio that appears on the right-hand side: SIR_i can be either the target SIR for mobile i , or it can be the actual SIR for mobile i , which is estimated from (1). The latter leads to more robust behaviour in cases when the mobile does not obey power control commands from the base station. On the other hand, as the investigations in this section show, using the actual SIR results in the performance being affected by power control errors. For both alternatives, SIR estimation errors affect performance. Although performance is affected, note that average service differentiation is still achieved.

Figures 2(a) and (b) show the transmission rate as a function of frame number, for continuous and discrete rate values, when the target SIR is used in (10). Each graph displays the rate for two mobiles with weights 1 and 2. Figures 3(a) and (b) show the same results when the actual SIR is used. Observe that when the target SIR is used, Figures 2(a) and (b), convergence is reached very fast, as soon as the last mobile has entered the system. On the other hand, when the actual SIR is used, Figures 3(a) and (b), convergence takes longer.

Figures 4(a) and (b) show the rates as a function of frame number, in the case of imperfect power control. Observe in Figure 4(a) that imperfect power control has no effect when the target SIR is used in the rate allocation. On the other hand, Figure 4(b) shows that imperfect power control has an effect when the actual SIR is used in the rate allocation. This occurs because the transmission powers appear in (1), hence power control errors affect the actual SIR . Despite the rate variations in Figure 4(b), observe that average service differentiation is still achieved.

Figures 5(a) and (b) show the rates as a function of frame number, when SIR estimation errors occur. These errors affect the estimation of the target SIR , hence they affect rate allocations when the target SIR is used, Figure 5(a). Moreover, errors in the target SIR result in fluctuation of the transmitting powers, hence they also affect rate allocations when the actual SIR is used, Figure 5(b). Comparison of Figures 5(a) and (b) with Figure 4(b) shows that SIR estimation errors have a slightly larger effect on rate fluctuations, compared to power control errors.

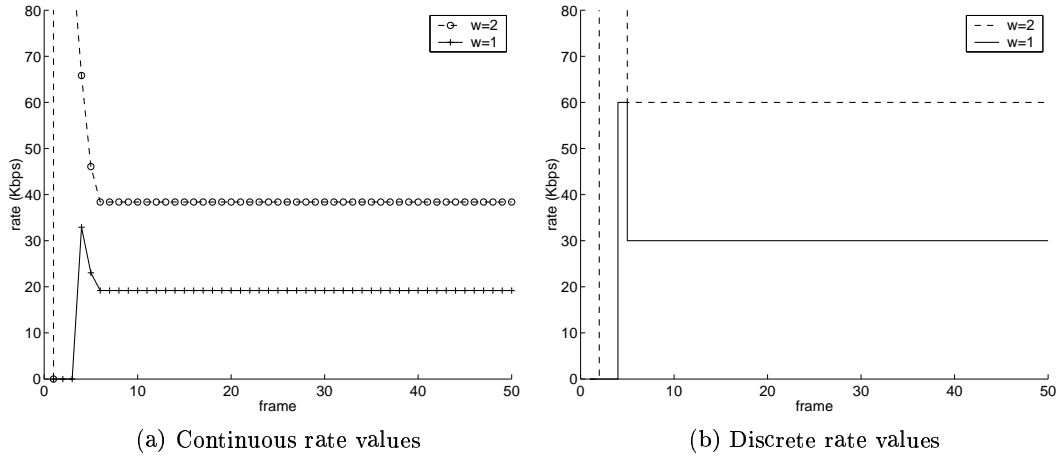


Figure 2: Rate as function of frame number, when rate allocation is based on the target SIR . $PCE = 0$, $SIR_{err} = 0$.

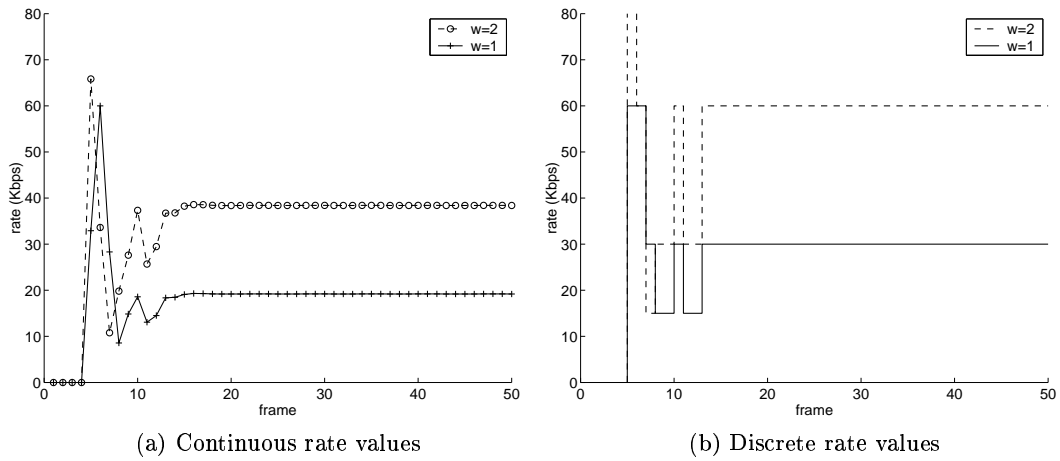


Figure 3: Rate as function of frame number, when rate allocation is based on the actual SIR . $PCE = 0$, $SIR_{err} = 0$.

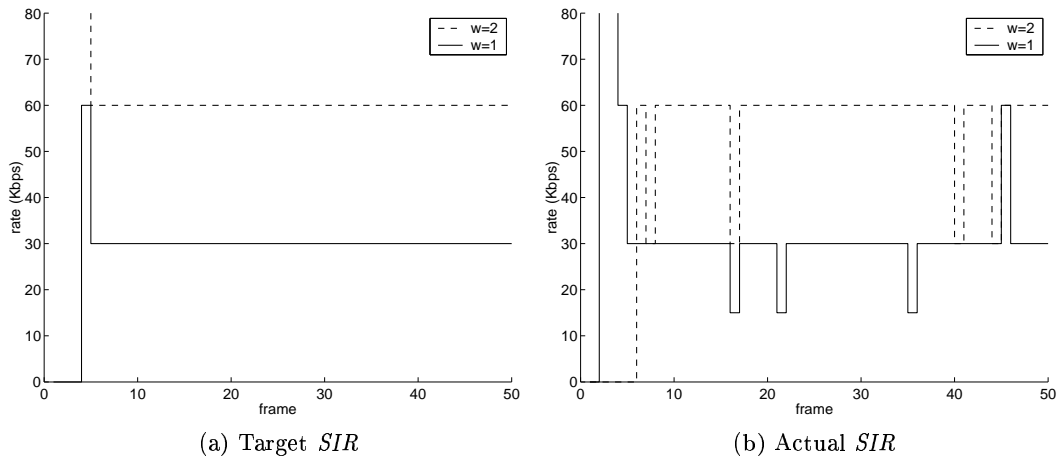


Figure 4: Rate (discrete values) as function of frame number. $PCE = 1$ dB, $SIR_{err} = 0$.

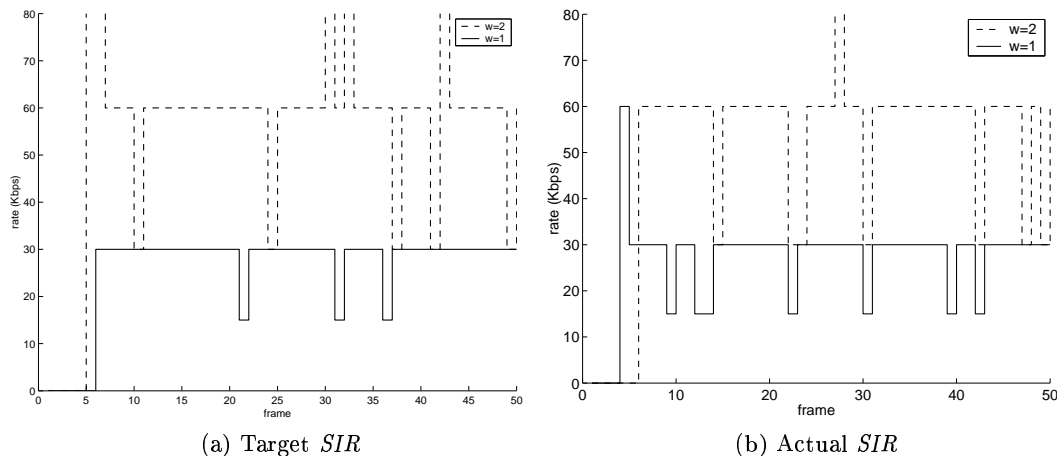


Figure 5: Rate (discrete values) as function of frame number. $PCE = 0$, $SIR_{err} = 1$ dB.

5 Conclusions

We have presented models and procedures for fair and efficient service differentiation in third generation mobile networks based on Wideband CDMA. With simulation experiments, we demonstrate the effectiveness of the procedures and investigate the effects of power control and SIR estimation errors.

Ongoing work includes extensions to the case of traffic that is both rate and signal quality adaptive and, unlike best-effort traffic, in addition to the average throughput of data transmission, is sensitive to losses. Related work involves the investigation of models and procedures for seamless congestion control in wireless and wired networks.

References

- [1] S. Dixit, Y. Guo, and Z. Antoniou. Resource management and quality of service in third-generation wireless networks. *IEEE Commun. Mag.*, pages 125–133, February 2001.
- [2] K. S. Gilhousen and *et al.* On the capacity of a cellular CDMA system. *IEEE Trans. on Vehicular Technology*, 40(2):303–312, May 1991.
- [3] Y. Guo and H. Chaskar. Class-based quality of service over air interfaces in 4G mobile networks. *IEEE Commun. Mag.*, pages 132–137, March 2002.
- [4] H. Holma and A. Toskala. *WCDMA for UMTS (revised edition)*. Wiley, New York, 2001.
- [5] L. Jorgueski, J. Farserotu, and R. Prasad. Radio resource allocation in third-generation mobile communication systems. *IEEE Commun. Mag.*, pages 117–123, February 2001.
- [6] S. Nanda, K. Balachandran, and S. Kumar. Adaptation techniques in wireless packet data services. *IEEE Commun. Mag.*, pages 54–64, January 2000.
- [7] A. Sampath, P. S. Kumar, and J. M. Holtzman. Power control and resource management for a multimedia CDMA wireless system. In *Proc. of IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, 1995.
- [8] V. A. Siris. Congestion pricing for resource control in Wideband CDMA. Technical Report No. 299, ICS-FORTH, December 2001.
- [9] L. C. Yun and D. G. Messerschmitt. Power control for variable QoS on a CDMA channel. In *Proc. of IEEE MILCOM'94*, NJ, USA, 1994.
- [10] J. Zander and S.-L. Kim. *Radio Resource Management for Wireless Networks*. Artech House, 2001.