

Performance Analysis of the Uplink of a CDMA Cell Supporting Elastic Services

Gábor Fodor¹ and Miklós Telek²

¹ Ericsson Research, SE-164 80 Stockholm, Sweden,
`Gabor.Fodor@ericsson.com`

² Budapest University of Technology and Economics, H-1111 Budapest, Hungary,
`telek@hit.bme.hu`

Abstract. We consider a single cell of a multi-service CDMA network, in which some of the service classes are explicit rate controlled. We call these elastic service classes. The instantaneous bit rate of elastic services is dynamically adjusted between a minimum and maximum value such that the system always remains work conserving. We develop a Markov model that allows us to study the impact of such state dependent (dynamic) rate control on the class-wise blocking probabilities and the first two moments of the holding times. We conclude that dynamic (state dependent) rate adjustment decreases the class-wise blocking probabilities and only moderately increases the expectation and the second moment of the time spent in the system.

1 Introduction

The Erlang capacity of code division multiple access (CDMA) systems has long been in the focus of research, basically starting ever since code division was first proposed as a multiple access technique; see for instance Chapter 6 of [1] and Section 5.7 of [2] which are dedicated to Erlang capacity issues in CDMA networks. Recently, the seminal paper by Altman [3], (see also [4]) showed that assuming perfect power control, slowing the transmission rates in the case of a single cell with heterogeneous QoS characteristics increases capacity. This result is non-trivial, since slowing the transmission rate leads to increased holding time of the session. Specifically, consider a multi-service system at which sessions arrive according to a Poisson process with intensity λ and assume that the session holding times are exponentially distributed with parameter μ . Then, assuming a *fixed* slow down rate a , the load (and thereby the Erlang capacity) of the multi-service system can increase from $\rho = \sum_i \lambda_i / \mu_i$ to $\rho_a > \rho$ ($a > 1$) under the same class-wise blocking probability constraints. The assumption on the system state independent fixed slow down rate a leads to a product form solution for the steady state that allows for a straightforward derivation of the blocking probabilities.

However, slowing down the transmission rates in all system states leads to a non-optimal resource utilization. Therefore, in this paper we propose a model that allows the instantaneous bit-rate of the elastic service classes to *adaptively* decrease/increase depending on the system state. The relaxation of the assumption that a is fixed is motivated by the observation that assuming perfect power control, the individual session bit-rates may change dynamically in the system depending on the currently on-going number of

Table 1. Model (Input) Parameters

| | |
|--------------|--|
| K | Number of service classes |
| $R_p(k)$ | Peak bit rate associated with class- k sessions |
| $\lambda(k)$ | Arrival intensity of sessions belonging to class- k |
| $1/\mu(k)$ | Mean (nominal) holding time of sessions belonging to class- k |
| $\hat{a}(k)$ | Maximum slow down (using the terminology of [3]) of $R_p(k)$ |
| i | Other cell (sector) interference to own cell (sector) interference ratio |
| $E(k)/N_0$ | Normalized signal energy per bit requirement of class- k |

sessions and their service class requirements. We intuitively expect that slowing down the session rates only in those system states in which it is necessary (as opposed to slowing down the elastic sessions in *all* system states) will lead to lower blocking probabilities and higher resource utilization. As we shall see, this is indeed the case. Unfortunately our setting destroys the product form solution but leads to a quasi birth-death structure that still allows for efficient solution of the system, at least when the number of service classes is low.

It has been observed by several papers that for elastic services not only the first, but also the second moment of the actual holding time has an impact on the user perceived QoS [5], [6]. Therefore, we also develop a modified Markov model that allows us to arrive at numerical results for the second moment of the elastic class holding time.

The paper is structured as follows. In the next Section we formulate the model and state the analysis objective in terms of performance measures of interest. Next, Section 3 examines the system state space and develops a solution for the steady state, from which various performance measures are derived. Subsequently, in Section 4, we describe the elements and the solution of the Markov model that allows us to derive the higher moments of the time spent in the system by elastic flows. Section 6 draws conclusions.

2 Model Assumptions and Performance Analysis Objective

2.1 The Input Parameters

We consider a single CDMA cell at which sessions belonging to one of K service classes arrive according to a Poisson arrival process of intensity $\lambda(k)$ ($k = 1, \dots, K$). Each class is characterized by a peak bit-rate requirement $R_p(k)$ and an exponentially distributed nominal holding time with parameter $\mu(k)$. When sending with the peak rate for a session, the required target ratio of the received power from the mobile terminal to the total interference energy at the base station is calculated as follows:

$$\tilde{\Delta}(k) = \frac{E(k)}{WN_0} \cdot R_p(k), \quad k = 1, \dots, K,$$

where $E(k)/N_0$ is the signal energy per bit divided by the noise spectral density that is required to meet a predefined QoS (e.g. bit error rate, BER); noise includes both

thermal noise and interference. This required $E(k)/N_0$ can be derived from link level simulations and from measurements. $R_p(k)$ is the peak bit rate of the session of class k and W is the spread spectrum bandwidth.

Just like in [3], let $M(k)$ be the number of ongoing sessions of class k . We will refer to vector $\mathbf{M} = \{M(k)\}$, $k = 1, \dots, K$ as the *state* of the system. We now assume that arriving sessions are blocked by a suitable admission control algorithm that prevents the system to reach the state in which the power that should be received at the base station would go to infinity. In other words, a suitable admission control algorithm must prevent the system to reach its *pole capacity* (as defined by Equation (8.10) of [7] and (5) of [3]).

In order to model this, we use the standard equations (8.3)-(8.12) from [7] as follows.

The power $P(k)$ that is received at the base station from the mobile terminal for session k must fulfill (assuming that the terminal can control the power level for each session separately):

$$\frac{P(k)}{P_N + I_{own} + I_{other} - P(k)} = \tilde{\Delta}(k), \quad k = 1, \dots, K \quad (1)$$

where I_{own} is the total power received by the base station within its cell (or sector), and I_{other} is the total power received from other cells (or sector). P_N is the background noise power. That is, $I_{own} = \sum_{k=1}^K M(k)P(k)$ and $I_{other} = i \cdot I_{own}$.

Rewriting (1), we get:

$$\frac{P(k)}{P_N + I_{own} + I_{other}} = \Delta(k), \quad \Delta(k) = \frac{\tilde{\Delta}(k)}{1 + \tilde{\Delta}(k)}, \quad k = 1, \dots, K \quad (2)$$

From (2):

$$\frac{P(k)}{P_N + (1+i) \sum_{l=1}^K M(l)P(l)} = \Delta(k), \quad k = 1, \dots, K$$

Multiplying with $M(k)$ and summing over k :

$$\sum_{k=1}^K P(k)M(k) = \frac{P_N \sum_{l=1}^K M(l)\Delta(l)}{1 - (1+i) \sum_{l=1}^K M(l)\Delta(l)} \quad (3)$$

From (3):

$$P(k) = \left(P_N + (1+i) \sum_{l=1}^K M(l)P(l) \right) \cdot \Delta(k), \quad k = 1, \dots, K \quad (4)$$

Substituting (3) into (4) and denoting $\Omega = \sum_{l=1}^K M(l) \cdot \Delta(l)$:

$$P(k) = \left(P_N + (1+i) \cdot \frac{P_N \cdot \Omega}{1 - (1+i) \cdot \Omega} \right) \cdot \Delta(k) = \frac{P_N \cdot \Delta(k)}{1 - (1+i) \cdot \Omega}, \quad k = 1, \dots, K \quad (5)$$

From (5) it is clear that the admission control algorithm must prevent that Ω reaches $1/(1+i)$. In the single class case, it means that the number of admitted sessions must fulfill: $M < \lfloor \Omega/\Delta \rfloor$ (where we now let $M = M(1)$ and $\Delta = \Delta(1)$). In the rest of the paper we assume that such admission control algorithm operates in the system. We note that because of the relation expressed by Equation (2) and the definition of Ω , one can consider Ω as the overall resource in the multi-rate CDMA system that the sessions must share. This observation can be seen as an analogy between the multi-rate CDMA model and the multi-rate loss models developed in the 80's and 90's [8]. The major difference between the classical loss models and the present CDMA model is that the relation between the slow down rate $a(k)$ and the resource consumption $\Delta_a(k)$ is not linear, as we shall see it in the next subsection and as it is also reflected by Equation (7).

2.2 The Impact of Slow Down

Recall that the required target ratio ($\Delta(k)$) depends on the required bit-rate. Explicit rate controlled elastic services tolerate a certain slow down of their peak bit-rate ($R_p(k)$) as long as the actual instantaneous bit rate remains greater than the minimum required $R_p(k)/\hat{a}(k)$. When the bit rate of a class- k session is slowed down to $R_p(k)/a(k)$, ($0 < a(k) \leq \hat{a}(k)$) its required $\Delta_a(k)$ value becomes:

$$\Delta_a(k) = \frac{\tilde{\Delta}(k)}{a(k) + \tilde{\Delta}(k)} = \frac{\Delta(k)}{a(k) \cdot (1 - \Delta(k)) + \Delta(k)}, \quad k = 1, \dots, K \quad (6)$$

which increases the number of sessions that can be admitted into the system, since now Ω_a must be kept below $1/(1+i)$, where

$$\Omega_a = \sum_{k=1}^K M(k) \cdot \Delta_a(k).$$

We use the notation $\Delta_{min}(k) = \Delta_{\hat{a}}(k)$ to denote the class-wise minimum target ratios (can be seen as the minimum resource requirement), that is when the session bit-rates of class- k are slowed down to that class' minimum value. It is the task of the bandwidth sharing policy to determine the $\Delta_a(k) \geq \Delta_{min}(k)$ values (and consequently the $a(k) \leq \hat{a}(k)$ class-wise instantaneous slow down factors) for each state of the system such that $\Omega_a < 1/(1+i)$. Because of the admission control assumption, such a resource assignment is always possible in feasible states.

2.3 Performance Measures

In order to define the performance measures of interest in this system, we make the following considerations. It is intuitively clear that the residency time of the elastic flows in this system depends not only on the amount of data they want to transmit (which is a random variable), but also on the instantaneous bit-rate they receive during their holding times. In order to specify this relationship we define the following quantities:

- $\theta(t, k) = R_a(t)$ defines the *instantaneous throughput* of a class- k elastic flow at time t . Note that $\theta(t, k)$ is a discrete random variable for any $t \geq 0$.
- $T_x(k) = \inf\{t \mid \int_0^t \theta(\tau, k) d\tau \geq x\}$ (random variable) gives the time it takes for the system to transmit x amount of data through a class- k elastic flow,
- $\hat{\theta}_x(k) = x/T_x(k)$ defines the *conditional throughput* of the class- k elastic flow during the transmission of x data unit. Note that $\hat{\theta}_x(k)$ is a continuous random variable.
- $\hat{\theta}(k) = \int_0^\infty \hat{\theta}_x(k) dG(x) = \mu(k)/R_p(k) \int_0^\infty \hat{\theta}_x(k) e^{-x \mu(k)/R_p(k)} dx$ (random variable) defines the *throughput* of the class- k elastic flow, where the amount of transmitted data is exponentially distributed with parameter $\mu(k)/R_p(k)$.

We note that from a user's perspective, it is in fact not the throughput, but the time it takes to transfer a file (image, email, etc) that has an impact on the perceived QoS. Therefore, in the numerical section we will focus on this performance measure.

In addition, we are interested in finding the blocking probabilities for all service classes i.e., $B(k), k = 1, \dots, K$.

The instantaneous and average *utilization* of the system are defined as follows:

$$U(t) = \sum_{k=1}^K M(t, k) \cdot \Delta_{a(t)}(k); \quad E[U(t)] = \int_0^\infty \tau dF_U(\tau),$$

where $F_U(\tau)$ denotes the stationary distribution of $U(t)$. In the above definitions it is emphasized that both $M(k)$ and $a(k)$ evolve in time.

3 Determining the Steady State of the System

3.1 Determining the System State Space

The maximum number of sessions from each class can be calculated as follows:

$$M_{max}(k) = \lfloor (\Delta_{\hat{a}(k)})^{-1} \rfloor, \quad k = 1, \dots, K$$

Recall that in each M state of the system, the inequality $\sum_k M(k) \cdot \Delta_a(k) < 1/(1+i)$ must hold. The states that satisfies this inequality are the *feasible states* and constitute the state space of the system which we will denote by Γ . The feasible states, in which the acceptance of an additional class- k session would result in a state outside of the state space are the class- k *blocking states*. The set of the blocking states is denoted by Γ^B . Due to the "Poisson Arrivals See Time Averages" (PASTA) property, the sum of

the class- k blocking state probabilities gives the (overall) class- k blocking probability [8].

In each feasible state, it is the task of the bandwidth sharing policy to determine the $\Delta_a(k) = f(M)$ values for each class. From these, the class-wise slow down factors and the instantaneous bit-rates of the individual sessions can be calculated as follows:

$$a(k) = \frac{\Delta(k) \cdot (1 - \Delta_a(k))}{\Delta_a(k) \cdot (1 - \Delta(k))}; \quad R_a(k) = R_p(k)/a(k) \quad (7)$$

3.2 Determining the Generator Matrix

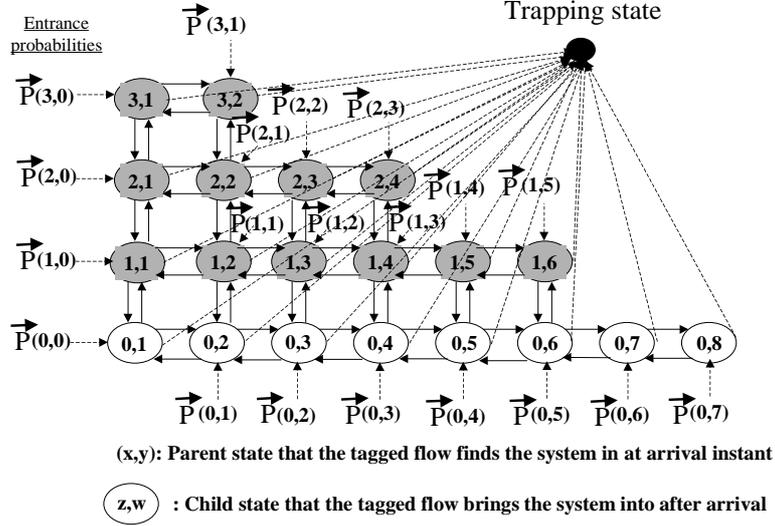


Fig. 1. Modified State Space with Trapping State

For ease of presentation, but without losing generality, we use the example illustrated by Figure 1 to develop the generator matrix of the state space. Assume that $i = 0$, $\hat{a}(1) = 1$, $\hat{a}(2) > 1$ and $\hat{a}(3) > 1$. In this case, the task of the bandwidth sharing policy simplifies to determining $\Delta_a(2)$ and $\Delta_a(3)$ for each state, from which $\hat{a}(2)$ and $\hat{a}(3)$ follows immediately.

We now make use of the assumptions that the arrival processes is Poisson and the nominal holding times are exponential. The system under these assumptions is a continuous time Markov chain (CTMC) whose state is uniquely characterized by the state vector M . In order to arrive at the performance measures of interest, we need to determine the CTMC's generator matrix Q and its steady state solution P .

Based on the considerations of the preceding subsections, we see that the generator matrix \mathbf{Q} possesses a nice structure, because only transitions between "neighboring states" are allowed in the following sense. Let $q(m_1, m_2, m_3 \rightarrow m'_1, m'_2, m'_3)$ denote the transition rate from state (m_1, m_2, m_3) to state (m'_1, m'_2, m'_3) . Then the non-zero transition rates between the feasible states are:

$$\begin{aligned} q(m_1, m_2, m_3 \rightarrow m_1 + 1, m_2, m_3) &= \lambda_1 \\ q(m_1, m_2, m_3 \rightarrow m_1, m_2 + 1, m_3) &= \lambda_2 \\ q(m_1, m_2, m_3 \rightarrow m_1, m_2, m_3 + 1) &= \lambda_3 \\ q(m_1, m_2, m_3 \rightarrow m_1 - 1, m_2, m_3) &= m_1 \cdot \mu_1 \\ q(m_1, m_2, m_3 \rightarrow m_1, m_2 - 1, m_3) &= m_2 \cdot \mu_2 / a_2 \\ q(m_1, m_2, m_3 \rightarrow m_1, m_2, m_3 - 1) &= m_3 \cdot \mu_3 / a_3 \end{aligned}$$

The first three equations represent the state transitions due to session arrivals, while the second three equations represent the transitions due to session departures. Here we utilized the fact that class-1 sessions cannot be slowed down, while class-2 and class-3 sessions can be slowed with a maximum of \hat{a}_2 and \hat{a}_3 factor respectively.

Note that the derivation of the generator matrix relies on the fact that the system is Markovian. This is not trivial because one could intuitively argue that since the elastic flows bring with themselves a certain amount of workload, the memoryless property does not hold, even if this workload is exponentially distributed. However, the Markovian property for such systems was independently of one another observed and formally proven by Altman *et al.* [9], Andersen *et al.* [5] and Nunez Queija *et al.* [10].

3.3 Deriving the Performance Measures

With the generator matrix in hand, we now need to solve the equation $\mathbf{P} \cdot \mathbf{Q} = 0$ (taking into account that $\mathbf{P} \cdot \mathbf{e} = 1$, where $\mathbf{e} = (1, \dots, 1)$). From the steady state distribution $\mathbf{P}(\mathbf{M})$, the performance measures of interest (except the second moments) immediately follow.

The blocking probabilities and the mean number of class- k sessions in the system are straightforward to derive (recall that $M(k) = \mathbf{M}[k]$):

$$B(k) = \sum_{\mathbf{M} \in \Gamma^B} \mathbf{P}(\mathbf{M}), \quad E[M(k)] = \sum_{\mathbf{M} \in \Gamma} M(k) \cdot \mathbf{P}(\mathbf{M}).$$

From this and Little's theorem we obtain the mean time that a class- k session spends in the system and the average resource utilization:

$$T(k) = \frac{E[M(k)]}{\lambda(k) \cdot (1 - B(k))}, \quad E[U] = \sum_{\mathbf{M} \in \Gamma} \sum_{k=1}^K M(k) \cdot \Delta_a(k).$$

From the steady state, the average class-wise throughput also follows:

$$\hat{\theta}(k) = \frac{\sum_{\mathbf{M} \in \Gamma} M(k) \cdot \mathbf{P}(\mathbf{M}) \cdot (1/a(\mathbf{M}))}{\sum_{\mathbf{M}} M(k) \cdot \mathbf{P} \cdot (\mathbf{M})}, \quad (8)$$

which provides an easy way to check and verify results, since the normalized throughput (to $R_p(k)$) is the reciprocal of the normalized mean holding time (to $1/\mu(k)$), which is indeed the case as we will see in the numerical section.

4 Solution Approach Based on a Markov Model with a Trapping State

The higher moments of the time spent in the system by the elastic sessions requires additional effort, which is the topic of this section.

4.1 Session Tagging and Modifying the State Space

We will continue to think of an elastic session as one that brings with itself an exponentially distributed amount of work and, if admitted into the system, stays in the system until this amount of work is completed. The method we follow here is based on (1) *tagging* an elastic session arriving to the system, which, at the time of arrival is in one of the feasible states; and (2) carefully examining the possible transitions from the moment this tagged call enters the system until it acquires the required service and leaves the system. Finally, un-conditioning on all possible entrance state probabilities, and applying results from [12], the moments of the best effort service time can be determined.

In this modified state space, we also define a *trapping (absorbing) state*, which corresponds to the state where the tagged session has acquired the requested amount of service and leaves the system. In this expanded state space the time until absorption [12] corresponds to the time the tagged session spends in the system. Indexing the modified state space in a similar manner as the original state space in Section 3, the new generator matrix, \tilde{Q} , will have the following structure:

$$\tilde{Q} = \begin{bmatrix} B & T \\ 0 & w \end{bmatrix} \quad (9)$$

where the B matrix represents the transitions between the non-trapping states, the T vector contains the transitions *to* the trapping state, the 0 vector indicates that no transitions are allowed *from* the trapping state, and $w = 0$. Once the structure of the expanded state space and the associated transition rates together with the initial probability vector, $P_R(0)$, are determined, we can apply the result of [13] (proved also in [14]) for the determination of the r^{th} moment of T_x :

$$T^{(r)} = r! \cdot P_R^T(0) \cdot (-B)^{-r} \cdot e \quad (10)$$

and specifically for the mean:

$$E[T] = P_R^T(0) \cdot (-B)^{-1} \cdot e, \quad (11)$$

which provides a natural way to check the results on the mean holding times numerically that are derived from the steady state as described in Section 3.

5 Numerical Results

5.1 Implementation

We have implemented the method described in Section 3 and Section 4 in a *Mathematica* script [15]. It takes the input parameters as described in Section 2 and generates the

Table 2. Input Parameters for Case I and Case II

| Parameter | Case I | Case II |
|--------------|-----------------------|-------------------|
| $\Delta(1)$ | $2 \cdot \Delta$ | $4 \cdot \Delta$ |
| $\Delta(2)$ | $3 \cdot \Delta$ | $4 \cdot \Delta$ |
| $\Delta(3)$ | $3 \cdot \Delta$ | Δ |
| $\hat{a}(1)$ | 1 | 1 |
| $\hat{a}(2)$ | 3 | 3 |
| $\hat{a}(3)$ | 1...6 | 1...6 |
| $\lambda(1)$ | $2 \cdot \lambda$ | λ |
| $\lambda(2)$ | $1.333 \cdot \lambda$ | λ |
| $\lambda(3)$ | $1.333 \cdot \lambda$ | $4 \cdot \lambda$ |

Table 3. Common Parameters for Case I and Case II

| Parameter | Value |
|----------------------------|-------|
| Δ | 0.049 |
| λ | 29.16 |
| $\mu(1) = \mu(2) = \mu(3)$ | 32.03 |

original and the modified state space and the corresponding generator matrixes. On a 750 MHz *Compaq Armada* PC with 256 MB RAM, this script (using Mathematica's *LinearSolve* function) is able to solve systems of around 4000 states in about 15 minutes.

5.2 Input Parameters

The input parameters for the two example cases that we study are summarized in Tables 3 and 2. We let $i = 0$. Class-1 is a "rigid" class in the sense that class-1 sessions cannot be slowed down once they admitted into the system. The maximum slow down value for class-2 sessions is kept fixed at $\hat{a}(2) = 3$, while we vary $\hat{a}(3) = 1 \dots 6$. In Case I, the class-wise load is: $\rho(k) = (\lambda(k)/\mu(k)) \cdot \Delta(k) = 4 \cdot \frac{\lambda \cdot \Delta}{\mu}$ and note that $\Delta(2) = \Delta(3)$, that is the (peak) resource requirement of the two elastic classes are the same. In Case II, the class-wise load is the same as in Case I, but now $\Delta(2) = 4 \cdot \Delta(3)$.

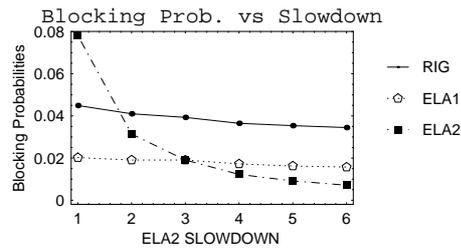


Fig. 2. Case I, Blocking Probabilities at Fixed Slow Down Rates: $a(2) = 3, a(3) = 1 \dots 6$

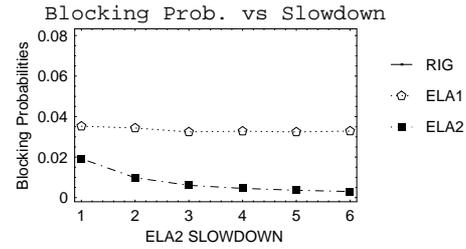


Fig. 3. Case 4, Case II, Blocking Probabilities at Fixed Slow Down Rates: $a(2) = 3, a(3) = 1 \dots 6$

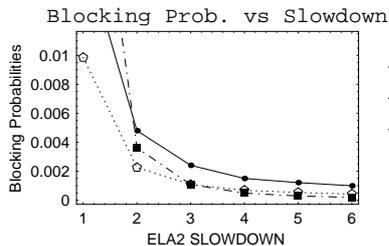


Fig. 4. Case I, Blocking Probabilities at State Dependent Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$

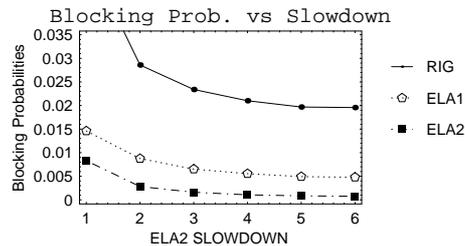


Fig. 5. Case II, Blocking Probabilities at State Dependent Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$

5.3 Blocking Probabilities

Figures 2 and 3 depict the class-wise blocking probabilities when the slow down rates are kept fixed as in [3] for Case I and II respectively. This fixed slow down rate is 3 for class-2 ("ELA1") and is set to $1 \dots 6$ for class-3 ("ELA2"). In Case I, where resource requirement of the two elastic classes are equal ($\Delta(2) = \Delta(3)$), the blocking probabilities are of course also equal when the class-3 slow down rate is also set to 3. More interestingly, we see that in Case I, allowing for a higher slow down value helps to reduce the class-3 blocking probability from 8% to around 1%. The class-1 blocking probabilities are not much affected and stay around above 4% and around 12% in the two cases.

It is noteworthy, that just as expected, the blocking rates for the system with dynamic slow down are lower than with fixed slow down. (Compare Figure 4 with Figure 2 and Figure 5 and with Figure 3.) This is of course because in the dynamic slow down case the sessions are not slowed down in the lightly loaded system states. Note also that in the fixed slow down system there is no need for a resource sharing policy, since the slow down rates are kept fixed in all system states.

5.4 Mean Time in System

Figure 6 and Figure 7 show the impact of slow down on the mean time spent by elastic flows in the system in Case I and Case II respectively. In Case I, as expected, the mean times are equal when $\hat{a}(3) = \hat{a}(2) = 3$. In Case II, class-2 sessions need to transmit significantly more data than class-3 sessions, since $R_p(2) > R_p(3)$. This clarifies that class-2 sessions spend somewhat more time in the system for all $\hat{a}(3)$ values.

In Figure 6 we notice that when $\hat{a}(3)$ is set to 2, the mean time for class-2 calls increases somewhat (from roughly 4% to 5%) as compared to the original holding time $1/\mu(2)$, while the blocking probabilities decrease significantly. When $\hat{a}(3)$ is further increased, the class-2 average time spent in the system decreases! This seemingly strange behavior is due to the fact that at $\hat{a}(3) = 2$, $\hat{a}(2)$ is still greater, and when class-3 sessions make better utilization of the resources, it is at the cost of class-2 resources. As class-3 sessions become "more and more elastic", it is the class-3 sessions that start spending more time in the system.

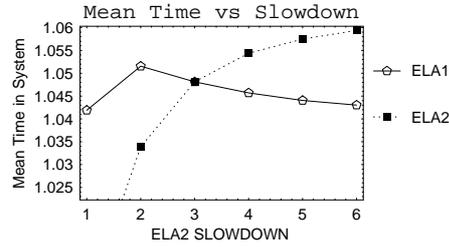


Fig. 6. Case I, Mean Time Spent in the System at *State Dependent* Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$ Normalized to $1/\mu$.

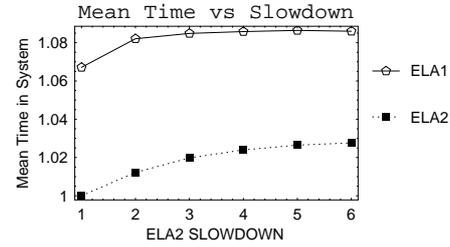


Fig. 7. Case II, Mean Time Spent in the System at *State Dependent* Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$ Normalized to $1/\mu$.

5.5 Second Moment of the Time Spent in the System

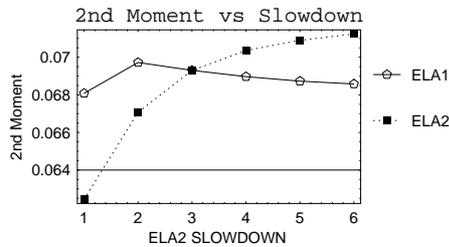


Fig. 8. Case I, Second Moment of the Time Spent in the System at *State Dependent* Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$ Normalized to $1/\mu$.

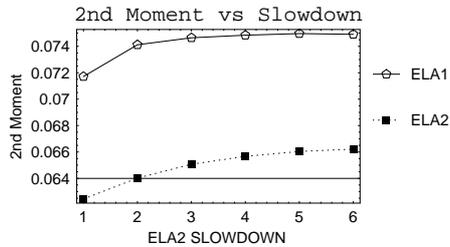


Fig. 9. Case II, Second Moment of the Time Spent in the System at *State Dependent* Slow Down Rates: $\hat{a}(2) = 3, \hat{a}(3) = 1 \dots 6$ Normalized to $1/\mu$.

The second moments show similar behavior as the mean time spent in the system (Figures 8 and 9.) We also see that increasing the maximum slow down factor beyond relatively small values (i.e. beyond $\hat{a}(3) = 4$) have only a small impact as compared to the impact when setting $\hat{a}(3) = 3$.

6 Conclusions

In this paper we considered an uplink capacity limited CDMA cell, where multiple service classes receive service. Service classes in this model are characterized by their Poisson arrival rates and the mean value of their exponentially distributed nominal holding times. So called rigid service classes occupy a fix amount of resource when they are admitted into the system. Elastic class sessions rate and thereby instantaneously occupied resources are adaptively adjusted between some minimum and maximum value.

This rather general flow level model can be seen as an extension of Altman's rate controlled CDMA model that is described in [3].

This model is analyzed in the paper by means of developing a Markov model and solving for the steady state. As it could be expected, the dynamic adjustment of the slow down rates result in significantly smaller blocking probabilities than that of the system of [3]. That is, the dynamic rate controlling system is expected to have higher Erlang capacity under the same blocking probability constraints.

References

1. A. J. Viterbi, "CDMA - Principles of Spread Spectrum Communication", *Addison-Wesley*, ISBN 0-201-63374-4, 1995.
2. S. Glisic, B. Vucetic, "Spread Spectrum CDMA Systems for Wireless Communications", *Artech House Publishing*, ISBN 0-89006-858-5, 1997.
3. E. Altman, "Capacity of Multi-service Cellular Networks with Transmission-Rate Control: A Queueing Analysis", *ACM Mobicom '02*, Atlanta, GA, September 23-28, 2002.
4. E. Altman, "Rate Control and QoS-related Capacity in Wireless Communications", - Keynote Speech at *Quality of Future Internet Services - QoFIS*, Stockholm, October 2003.
5. A. T. Andersen, S. Blaabjerg, G. Fodor, M. Telek, "A Partially-Blocking Queueing System with CBR and ABR Arrival Streams" *Telecommunication Systems 19:1*, Kluwer Academic Publishers, pp. 75-99, 2002.
6. S. Racz, B. P. Gero, G. Fodor, "Flow Level Performance Analysis of a Multi-service System Supporting Elastic and Adaptive Services", *Performance Evaluation 49*, Elsevier, pp. 451-469, 2002.
7. H. Holma, A. Toskala, "WCDMA for UMTS - Radio Access for Third Generation Mobile Communications", Wiley, ISBN 0 471 72051 8, First Edition, 2000.
8. K. W. Ross, "Multiservice Loss Models for Broadband Telecommunication Networks", ISBN 3-540-19918-8, *Springer Verlag*, 1995.
9. Eitan Altman, Damien Artiges and Karim Traore, "On the Integration of Best-Effort and Guaranteed Performance Services", *INRIA Research Report No. 3222*, July, 1997.
10. R. Nunez Queija, J. L. van den Berg, M. R. H. Mandjes, "Performance Evaluation of Strategies for Integration of Elastic and Stream Traffic", *International Teletraffic Congress*, UK, 1999.
11. L. Massoulié and J. Roberts, "Bandwidth Sharing and Admission Control for Elastic Traffic", Available at: <http://www-sop.inria.fr/mistral/pub/massoulie.html>.
12. R. A. Howard, "Dynamic Probabilistic Systems and Markov Decision Processes", *MIT Press*, 1960.
13. J. A. Buzacott, "Markov Approach to Finding Failure Times of Repairable Systems", *IEEE Transactions on Reliability*, Vol. R19, pp. 152-156, November 1970.
14. B. F. Nielsen, "Modeling Multiple Access Systems with Phase Type Distributions", Ph.D. Thesis, *IMSOR, Technical University of Denmark*, Thesis No. 49, 1988.
15. S. Wolfram, "The Mathematica Book", V4.0, ISBN 0521643147, 1999.