Performance analysis of link carrying capacity in CDMA systems*

Notker Gerlich[a], Phuoc Tran-Gia[a], Khaled Elsayed[b], and Nikhil Jain[b]

[a]Institute of Computer Science, University of Würzburg, Am Hubland, D-97074 Würzburg, FRG.

[b]Wireless Networks, Northern Telecom, Richardson, TX 75083-3871, USA.

Code Division Multiple Access (CDMA) technology is gaining momentum as the preferred wireless system for the next generation Personal Communication Systems (PCS). In CDMA systems, voice is encoded and packaged in variable length packets that are transported between the mobile station and the switching center. Although the packetization provides a great flexibility in resource allocation, it poses a Quality of Service (QoS) problem on voice. In this paper, we discuss link dimensioning for a typical CDMA encoder. We consider a T1/E1 link extending between a CDMA basestation and the Mobile Switching Center. Traffic from various voice sources is subject to a framing scheme, which presents a semi-periodic batch input at the T1/E1 interface cards. We analyze the resulting queuing system using discrete-time analysis and verify our results by simulation. Our results show the accuracy of the analysis and the potential statistical gain that can be achieved by voice packetization.

1. INTRODUCTION

The Code Division Multiple Access (CDMA) cellular system promises many advantages over its AMPS and TDMA counterparts. The advantages include enhanced privacy, resistance to jamming, improved voice quality, improved handoff performance, and soft (and increased) capacity [10, 2]. One important aspect of CDMA is the usage of a variable bit rate (VBR) voice encoder which reduces the required bandwidth (on the land network) and interference (on the airlink). The vocoder detects speech and silence in the voice process and adjusts its rate accordingly. It also tunes out background noise and dynamically varies its data transmission rate to operate at one of four different levels.

The VBR vocoder has impact on the airlink interface capacity as shown in [11]. Though the air link capacity is the scarce resource in a cellular system, it is nonetheless important to optimize the usage and design the land interconnecting network efficiently. A base station (BS) is connected to a Mobile Switching Center (MSC) via leased lines. These leased lines are quite expensive and add to the cost of operating the cellular system.

In this paper we study the issue of BS-MSC interconnection. We provide a methodology for performance analysis and dimensioning of the involved communication links to satisfy the quality of service. The quality of service is expressed in terms of a delay bound

*Parts of this paper are based on research supported by the Deutsche Forschungsgemeinschaft (DFG) under grant Tr-257/3.

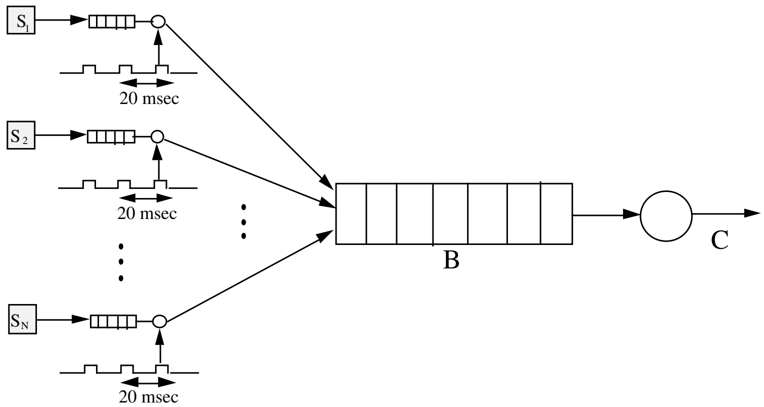

Figure 1: Voice Path from Mobiles to the BS–MSC link

(maximum allowable delay) and a packet loss probability. We use discrete-time analysis methodology as developed in [1, 5, 6] to analyze the problem and validate our results using simulation.

The rest of this paper is organized as follows. In Section 2, we describe the problem under consideration and state the objectives of studying it. In Section 3, the analysis of the model is presented. In Section 4, we provide a numerical study and validation of the models introduced. Section 5 concludes the paper.

2. PROBLEM DESCRIPTION

Consider a basestation in a CDMA system. The major functionality of the BS is to perform the IS-95 air interface specifics and provide connection between the mobile users and the central office switch. Let us focus our attention on the BS–MSC link in the reverse direction (i.e., from BS to MSC). Let the link capacity be C bits/sec. The capacity typically comes in multiples of the DS0 channel rate which is 64 *kbps* or 56 *kbps* (depending on line coding used). The link is statistically shared among both the packetized voice traffic of the connected sources and the signaling packets. A limited buffer of size B bits is provided to store the incoming packets until the link becomes available.

The link interface card implements a data link layer which is very similar to the HDLC protocol. Higher priority is given for signaling packets with regard to buffer sharing. Signaling packets can push out already existing voice packets if they find a full buffer.

Let us now describe the sequence of operations performed on voice packets until they reach the link buffer. At the mobile station, the vocoder adapts its rate according to speech activity, noise and threshold. In steady state, an 8K vocoder transmits in one of four rates with a probability as shown in Table 1.

Table 1: Rate distribution and corresponding packet lengths

Rate [*bps*]	Packet Length [*bit*]	Probability
9600	256	0.291
4800	160	0.039
2400	120	0.072
1200	96	0.598

Another version of the vocoder that provides better voice quality operates at a maximum rate of 14.4 *kbps* (13K vocoder). In this work, we focus on the 8K vocoder without loss of generality. At the base station, for each mobile station a digital signal processor unit (ASIC) is allocated that performs IS-95 processing and retrieves the packetized voice data from the raw IS-95 stream. A processor attached to the ASICs will schedule packet transmission from the different sources according to a framing mechanism described as follows. A system wide frame of $T = 20$ *msec* is used to multiplex the packets. The frame is divided into M slots with length T/M *msec*. For each voice source one slot is assigned during call setup and the source is only allowed to transmit packets in the assigned slot. This slot may change only when a call goes through a hard handoff. The slot assignment is done such that the load is distributed evenly among the slots. It is possible that more than one source is assigned to the same slot.

The link buffer receives the packetized voice and signaling (related to call processing) in its common buffer and transmits the packets in a FIFO manner.

The quality of service for the BS–MSC link is defined as follows:

- Maximum delay for an arbitrary packet should be less than d *msec*. Typically, d is set to 4 *msec*. Since the delays involved are random, we express the maximum delay as the 99.99% quantile of the delay distribution of an arbitrary packet.
- The packet loss probability due to the finite buffer should be kept below ϵ (where ϵ is typically $10^{-3} - 10^{-6}$).

When designing a system the following questions need to be addressed. What is the minimum link speed C required for a given number N of voice channels which the BS is serving? The other way round, what is the number N of voice channels that can be supported by a given link capacity C? Finally, what is the appropriate buffer size B to sustain the required quality of service? Large buffers would increase the maximum possible delay while enhancing the loss performance.

Signaling traffic is assumed to generate α% of the voice traffic. Typically α is about 1% to 10%. We concentrate on the voice traffic only and we can scale the obtained results to reflect the effect of signaling.

3. QUEUING MODEL AND SOLUTION

The problem described above can be abstracted as follows. Due to the framing structure the voice packet arrivals are a deterministic arrival process with a given number of arrivals at each slot. This is only true, however, for the limited period of time where we have a particular call mix. Since a very large number of slot allocations is possible considering all possible realizations results in a binomial distribution of the number of packets in a given slot. The service time is given by the packet length divided by link speed C.

We consider the bit buffer as a finite capacity queuing model operating in discrete time. Time is discretized into intervals of unit length Δ, which is the transmission time of a single data-unit. The size of a data-unit is given by the greatest common divisor of the packet lengths as given by a discrete r.v. (random variable) V (in this work the distribution of V is given by Table 1). During a single duty cycle of constant length a, which is a multiple of Δ, packets transmitted by the active fraction of N sources are collected and submitted to the buffer. Thus, we have arrivals of batches of packets, each of which is a batch of data-units. The number of active sources, and hence the number of packets in a batch, is governed by a binomially distributed r.v. $X \approx b(N, 1/16)$, where $b(z, p)$ denotes the binomial distribution with parameters z and p.

Two different admission policies are considered, if an arriving packet does not fit into the buffer:

1. *Partial packet loss*: free positions of the buffer are filled; the remaining data-units are lost.
2. *Full packet loss*: the packet is lost as a whole.

It should be noted that partial packet loss policy does not make sense from the implementation's point of view. Nevertheless, it makes sense in terms of the state analysis which is simplified for partial packet loss. For the parameter sets investigated here assuming partial packet loss leads to the same results as full packet loss as will be seen later.

Some related queuing problems have already been considered in the literature before. In [9] a single server infinite queue with Poisson batch arrivals and general service times ($M^{[X]}/G/1$) is studied. [12] derives the generating function of the state probabilities of the $GI^{[X]}/M/c$ system. Statistical multiplexers for packetized voice connections are investigated in [4] and [8]. All these models have infinite buffers in common whereas we have to deal with a finite buffer.

In [3] and [7] finite queuing systems with batch arrivals are investigated. The blocking probabilities for full and partial packet loss of the $M^{[X]}/M/1-S$ and $G^{[X]}/D/1-S$ queuing systems are derived, respectively. The difference to the queuing system investigated here is that in our case the batching process has two stages: batches of packets, each of which consists of a number of data-units.

3.1. State Analysis

3.1.1. Partial packet loss

When analyzing the buffer occupancy distribution using discrete time analysis technique [1, 5, 6] one keeps track of the time-dependent unfinished work process. This process is described by the r.v.'s

- U_n^- r.v. for the number of data-units present in the buffer immediately *prior* to the arrival instant of the nth batch;
- U_n^+ r.v. for the number of data-units in buffer immediately *after* the arrival instant of the nth batch;
- Y_n r.v. for the number of data-units in batch n.

The discrete distributions of these r.v.'s are denoted by $u_n^-(k)$, $u_n^+(k)$, and $y_n(k)$, respectively. The relation between these r.v.'s is given by

$$U_n^+ = \min\{U_n^- + Y_n,\ s\}, \tag{1a}$$

$$U_{n+1}^- = \max\{U_n^+ - a,\ 0\}. \tag{1b}$$

In terms of the discrete distributions the last equation reads

$$u_n^+(k) = \pi^s[\, u_n^-(k) \circledast y_n(k)\,] \tag{2a}$$

$$u_{n+1}^-(k) = \pi_0[\, u_n^+(k) \circledast \delta(k+a)\,], \tag{2b}$$

where $\pi^s[\cdot]$ and $\pi_0[\cdot]$ are linear sweep operators on probability distributions defined by

$$\pi^m[z(k)] = \begin{cases} z(k) & \text{for } k < m, \\ \displaystyle\sum_{i=m}^{\infty} z(i) & \text{for } k = m, \\ 0 & \text{for } k > m; \end{cases} \tag{3a}$$

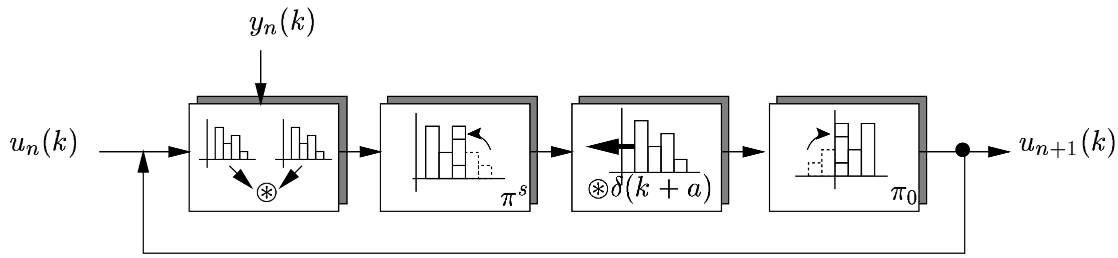

Figure 2: computational diagram for buffer occupancy distribution

$$\pi_m[z(k)] = \begin{cases} 0 & \text{for } k < m, \\ \sum_{i=-\infty}^{m} z(i) & \text{for } k = m, \\ z(k) & \text{for } k > m; \end{cases} \tag{3b}$$

and '$\circledast$' denotes the discrete convolution

$$z(k) = z_1(k) \circledast z_2(k) = \sum_{i=-\infty}^{+\infty} z_1(k-i) \cdot z_2(i). \tag{4}$$

Note, that the convolution of a distribution $z(k)$ and the distribution defined by the Kronecker-function

$$\delta(k+a) = \begin{cases} 1 & \text{for } k+a = 0, \\ 0 & \text{for } k+a \neq 0 \end{cases} \tag{5}$$

denotes a shift of $z(k)$ by a indices.

Since Eqn. (2) represents a recursive relation between the system states seen upon arrival by consecutive batches,

$$u^-_{n+1}(k) = \pi_0[\,\pi^s[u^-_n(k) \circledast y_n(k)] \circledast \delta(k+a)\,], \tag{6}$$

it gives rise to the algorithm depicted in the computational diagram Figure 2. It should be noted that the convolutions involved can be computed efficiently by using fast Fourier transforms (FFTs). In our case of identically and independently distributed batch sizes Y_n (in data-units) the computational diagram describes an iterative algorithm to determine the equilibrium buffer occupancy distribution

$$u^-(k) = \lim_{n\to\infty} u^-_n(k). \tag{7}$$

To complete the derivation we have to give the distribution $y_n(k)$ of r.v. Y_n. Since Y_n is the sum of X r.v.'s V, $y_n(k)$ is the compound distribution of the corresponding distributions $x(k)$ and $v(k)$:

$$y_n(k) = \sum_{i=0}^{N} v^{\circledast i}(k) \cdot x(i), \tag{8}$$

where, $v^{\circledast i}(k)$ denotes the i-fold convolution of $v(k)$ with itself and, naturally, $v^{\circledast 0}(k) = \delta(0)$.

3.1.2. Full packet loss

If we employ this policy the whole packet is lost if it does not fit into the buffer upon arrival. We obtain the following equations for the system-state r.v.'s as defined above

$$U_n^+ = \begin{cases} U_n^- + Y_n & \text{if} \quad U_n^- + Y_n \le s \\ U_n^- + \tilde{Y}_n & \text{if} \quad U_n^- + Y_n > s \end{cases} \tag{9a}$$

$$U_{n+1}^- = \max\{U_n^+ - a,\ 0\}, \tag{9b}$$

where r.v. $\tilde{Y}$ denotes the fraction of Y_n that is accepted. Distributions of these r.v.'s are given as

$$u_n^+(k) = [u_n^- \circledast y_n](k) \quad + \left[\sum_{i=s-k+1}^{\infty} v(i)\right] \cdot \sum_{j=1}^{N} x(j) \sum_{i=0}^{j-1} [u_n^- \circledast v^{\circledast i}](k) \qquad k = 0, 1 \ldots s, \tag{10a}$$

$$u_{n+1}^-(k) = \pi_0[\, u_n^+(k) \circledast \delta(k+a)\,], \tag{10b}$$

where Eqn. (10a) is derived as follows. Consider the arrival of a batch of j packets which occurs with probability $x(j)$. If the batch is small enough to fit into the buffer completely $U_n^+ = k$ if $U_n^- + Y_n = k$. The batch does not fit into the buffer and is truncated to leave $U_n^+ = k$ if for any $0 \le i \le j-1$ i packets together with U_n^- sum up to k and the length of the $(i+1)$th packet exceeds $s-k$. Unconditioning with respect to variables i and j gives rise to Eqn. (10a).

As with partial packet loss, Eqn. (10) represents a recursive relation between system states seen upon arrival of consequent batches and, hence, gives rise to an iterative algorithm to calculate the state probabilities in equilibrium.

3.2. Packet Loss Probability and Waiting Time Distribution

The derivation of the packet loss probability and the waiting time distribution applies for both partial and full packet loss policy.

Observing a tagged packet, we define the r.v. Y^* to be the end, i.e., the last data-unit of this packet within its batch of packets; $y^*(k)$ denotes the corresponding distribution. Clearly, given a buffer occupancy of U upon arrival of the batch the tagged packet is lost if $U + Y^* > s$. This leads to the loss probability as given by

$$p_{\text{loss}} = \sum_{i=s+1}^{\infty} [u \circledast y^*](i), \tag{11}$$

where distribution $y^*(k)$ remains to be derived.

To that end we define the conditional distribution $y^*_{|X=j}(k)$, which denotes the distribution of the tagged packet's end arriving within a batch of j packets. Since the position of the tagged packet within the batch is distributed uniformly by applying complete probability formula we obtain

$$y^*_{|X=j}(k) = \sum_{i=1}^{j} v^{\circledast i}(k) \cdot \frac{1}{j}. \tag{12}$$

The probability for the tagged packet to arrive within a batch of j packets is $j \cdot x(j)/E[\ X\ ]$, where the operator $E[\ Z\ ]$ denotes the expectation of r.v. Z. Thus, unconditioning with respect to j gives

$$\begin{aligned} y^*(k) &= \frac{1}{E[\ X\ ]} \sum_{j=1}^{N} y^*_{|X=j}(k) \cdot j \cdot x(j) \\ &= \frac{1}{E[\ X\ ]} \sum_{j=1}^{N} x(j) \sum_{i=1}^{j} v^{\circledast i}(k). \end{aligned} \tag{13}$$

The waiting time distribution $w(k)$ for packets transmitted is given by the buffer occupancy U that a tagged packet which is granted admission sees plus the work brought into the system by the packets ahead of it (including the tagged packet itself), Y^*. Defining thus

$$\tilde{U}^- = U^- + Y^* \tag{14a}$$

$$\tilde{u}^- = u^-(k) \circledast y^*(k) \tag{14b}$$

the waiting time distribution reads

$$w(k) = \begin{cases} \dfrac{\tilde{u}^-(k)}{\sum_{i=0}^{s} \tilde{u}^-(i)} & 0 \le k \le s, \\ 0 & k > s. \end{cases}$$

4. RESULTS

The quality of service for the BS–MSC link is defined in terms of both the delay of an arbitrary packet and the packet loss probability: The delay should be less than $d = 4$ *msec* for 99.99% of the packets and the packet loss should be below some $\epsilon \in [10^{-6}, 10^{-3}]$. Figures 3 and 4 show the probability for a packet to experience a delay of more than $d = 4$ *msec* and the packet loss probability versus the number of voice channels, respectively. The packet length distribution is the distribution of Table 1 and the buffer size is $B = 2$ *kB*. The diagrams clearly indicate that the limitation of the packet delay constitutes the stronger constraint. Take for instance the 16×64 *kbps* curve: the loss probability constraint suggests a support of 130 voice channels whereas the delay limitation allows only a number of 110 voice channels to be supported. Since both the loss and the delay curves are steep the quality of service improves significantly if one allows a few voice channels less to be supported. Consider again the 16×64 *kbps* curve: the loss probability drops 3 orders of magnitude when supporting only 130 channels instead of 135. In the following, we will review the questions issued in Section 2. Table 2 shows the maximum number of voice channels that can be supported given a certain link capacity (in multiples of 64 *kbps*). The resulting statistical multiplexing gain is depicted in Figure 5. The multiplexing gain is defined as the ratio of the maximum number of voice channels that can be supported and the minimum number of voice channels supported when applying peak bit rate allocation. The latter value is given by the ratio of the link capacity and the maximum voice channel bit rate. The statistical multiplexing provides a reasonable gain over peak rate allocation for rates larger than 8×64 *kbps* and the multiplexing gain reaches 1.5 at 30×64 *kbps*. But the multiplexing gain smaller than 1 for low channel rates cannot be interpreted as advantage of peak rate allocation over statistical multiplexing. With peak rate allocation

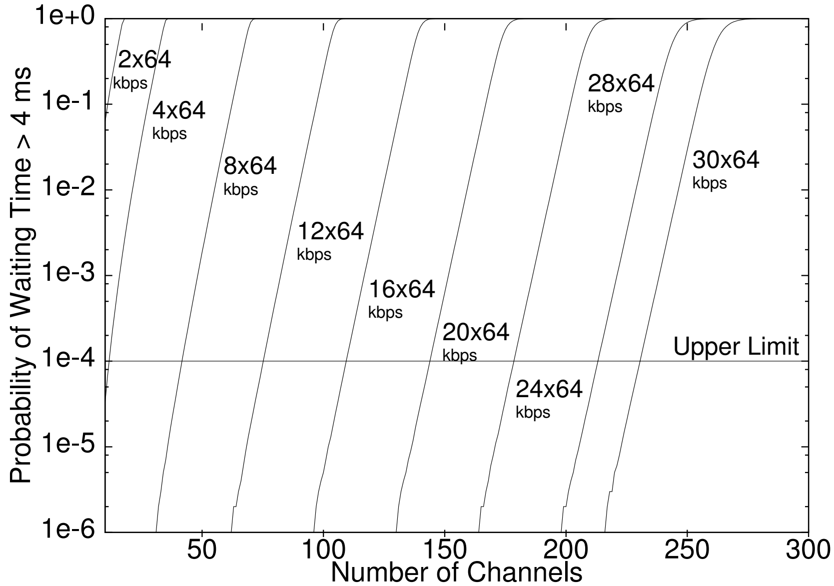

Figure 3: Complementary Packet Delay Distribution

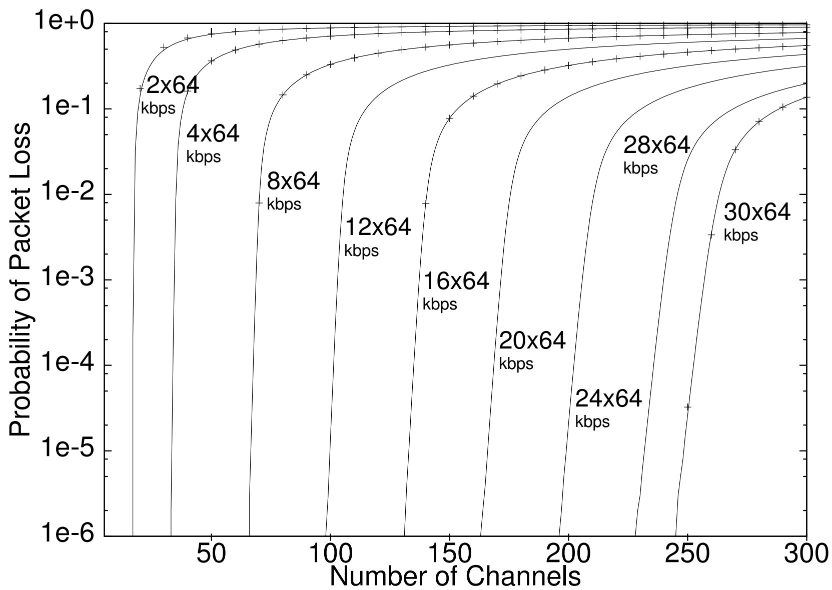

Figure 4: Packet Loss Probability

the calculation of the maximum number of supported voice channels only considers packet loss probability but not packet delay. To avoid exceeding the delay limit, all packets of a batch must be transmitted within 4 *msec*, which is not possible if the batch is large (i.e. the number of supported sources is large) and the channel rate is low.

Figure 6 illustrates the influence of the buffer size on the quality of service as expressed by the packet loss probability. The link speed is 30×64 *kbps* which is equivalent to 300 data-units per duty cycle. Consequently, a buffer smaller than 300 data-units is emptied in each duty cycle; each batch of packets finds an empty buffer. This explains the bends at $B = 300$ data-units for 200, 225, 250, and 300 channels. For buffers larger than 300 data-

Table 2: Maximum Number of Multiplexable Voice Channels

Link Speed [$64kbps$]	Voice Channels
2	2
4	11
8	41
12	75
16	109
20	143
24	178
28	213
30	230

units a queue builds up in the buffer and the typical exponential tail can be observed. The 100 channels curve exhibiting no bend is due to the fact that the length of the maximum possible batch is 320 data-units. A comparison between the curves shows that the buffer size required to guarantee a packet loss smaller than 10^{-6} increases non-linearly with the number of channels to support.

Finally, to show the accuracy of the results obtained, Figure 4 also shows simulation results for full packet loss for typical parameter sets. Since the crosses (+) are covering the curves, the results are accurate and one may use the part loss computation which is easier to implement and runs with less computational effort.

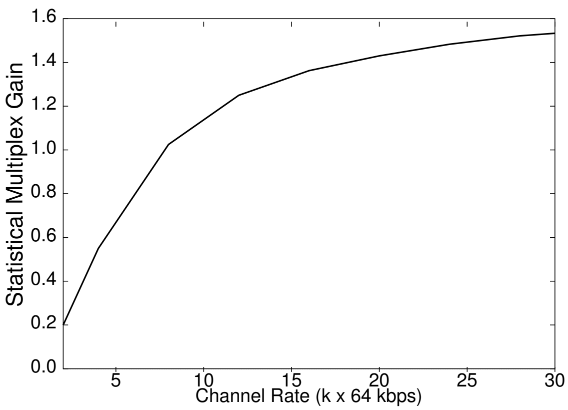

Figure 5: Statistical Multiplexing Gain

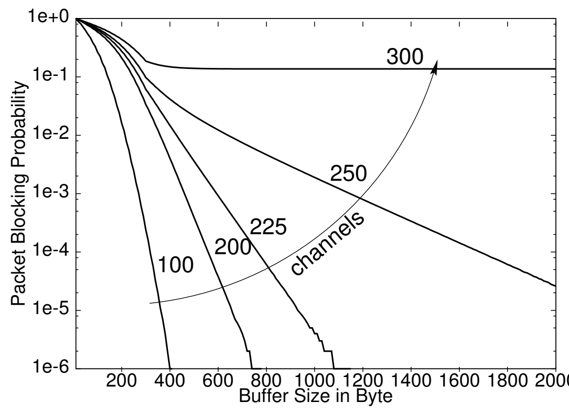

Figure 6: Buffer Size

5. CONCLUSION AND OUTLOOK

We presented a performace study of the statistical multiplexing of packetized voice in CDMA systems. The results from our model compare favorably with the simulation and show the accuracy of the approximation method. We have seen that for almost all practical link speeds, statistical multiplexing provides a reasonable gain over peak rate allocation (or circuit switching).

The model does not capture the correlation in the arrival process which is caused by the fact that the voice encoding process is dependent on voice activity detection. As an extension to this work, ongoing studies focus the assumption that voice sources are modeled by a multi-rate Markov modulated process. The system analysis is then decomposed into two phases: first, we capture the short term queue dynamics using the technique presented in this paper. Then we use large deviations theory to study the queue dynamics when the sources are multi-rate Markov modulated process. The results from the two phases can then be combined to provide a better understanding of the queuing behavior.

Acknowledgements
The programming efforts of B. Neuner are much appreciated.

REFERENCES

[1] M. H. Ackroyd. Computing the waiting time distribution for the G/G/1 queue by signal processing methods. *IEEE Transactions on Communications*, COM-28(1):52–58, January 1980.

[2] J. D. Gibson, editor. *The Mobile Communications Handbook*. IEEE Press / CRC Press, College Station, TX, 1996.

[3] D. R. Manfield and P. Tran-Gia. Analysis of a finite storage system with batch input arising out of message packetization. *IEEE Transactions on Communications*, COM-30(3):456–463, 1982.

[4] G. Ramamurthy and B. Sengupta. Delay analysis of a packet voice multiplexer by the $\sum D_i/D/1$ queue. *IEEE Transactions on Communications*, COM-39(7):1107–1114, 1991.

[5] P. Tran-Gia. Discrete-time analysis for the interdeparture distribution of GI/G/1 queues. In O. J. Boxma, J. W. Cohen, and H. C. Tijms, editors, *Teletraffic Analysis and Computer Performance Evaluation*, pages 341–357. North–Holland, Amsterdam, 1986.

[6] P. Tran-Gia. Discrete-time analysis technique and application to usage parameter control modelling in ATM systems. In *Proc. 8th Australian Teletraffic Seminar*, Melbourne, Australia, December 1993.

[7] P. Tran-Gia and H. Ahmadi. Analysis of a discrete-time $G^{[X]}/D/1-S$ queueing system with applications in packet–switching systems. In *Proc. INFOCOM '88*, pages 861–870, 1988.

[8] S. Tsakiridou and I. Stavrakakis. Mean delay analysis of a statistical multiplexer with batch arrivals processes — a generalization to Viterbi's formula. *Performance Evaluation*, 25:1–15, 1996.

[9] J. Van Ommeren. Simple approximations for the batch-arrival $M^X/G/1$ queue. *Operations Research*, 38(4):678–685, 1990.

[10] A. M. Viterbi. *CDMA Principles of Spread Spectrum Communication*. Addison–Wesley, Reading, MA, 1995.

[11] A. M. Viterbi and A. J. Viterbi. Erlang capacity of a power controlled CDMA system. *IEEE Journal on Selected Areas in Communications*, 11(6):892–899, August 1993.

[12] Y. Zhao. Analysis of the $GI^X/M/c$ model. *Queueing Systems*, 15:347–364, 1994.