MULTIUSER ACCESS CAPACITY OF PACKET SWITCHED CDMA SYSTEMS

Aylin Yener   Roy D. Yates
Wireless Information Network Laboratory (WINLAB), Rutgers University
yener@winlab.rutgers.edu   *ryates@winlab.rutgers.edu*

Abstract: We consider packet DS-CDMA systems with a connectionless architecture where no dedicated connection is kept between users and the system. Users have to access the system using the same signature sequence and have to be acquired each time they need to send packets. Due to the asynchronous nature of the system, multiple users can be accommodated even when a single signature is available. However, we observe that even under optimistic assumptions, the capacity of such a system is less than that of G orthogonal ALOHA channels that can be accommodated in the same bandwidth of this CDMA system with processing gain G.

I Introduction

Future wireless systems will require flexibility in terms of accommodating a variety of services with different requirements. Code Division Multiple Access (CDMA) technology emerges as suitable candidate for such a system architecture. For next generation CDMA systems, employment of packet based data communication services is anticipated [1]. Considerable effort has been directed towards the performance analysis and establishing efficient protocols for packet CDMA networks to date (see for example, [2] and references therein). All these studies dealt with issues after the timing parameters and the activity status of users are obtained. This is a valid assumption for systems where users have long and/or frequent communication sessions and thus can afford to first establish connection with the system and then keep this dedicated connection, essentially leading to a circuit switched architecture. However, for services with relatively short sessions, such as packet data, the aforementioned model is not appropriate. Instead, a system with no dedicated resources should be considered.

A packet switched CDMA system has a *connectionless* architecture, i.e. no connection is to be kept between a user and the base station except when the user is sending information. User recognition and synchronization has to be established every time a user needs to transmit information. Also, users do not have assigned signature sequences and have to send information using one of the predetermined signature sequences. Existing analyses on such systems often ignore the fact that user acquisition (user recognition and timing acquisition) has

to be achieved for every transmission or state that it can be achieved readily (e.g. [3] and more recently [4, 5]). On the other hand, for circuit switched CDMA there is a body of research that concentrates on user parameter acquisition (e.g. [6]) and it has been well known that timing acquisition can be capacity limiting [7].

In this paper, we investigate the capacity of packet switched CDMA systems taking into account the fact that user acquisition has to be achieved during each information transfer session, e.g. each packet or frame. The emphasis of our results is on the fact that acquisition and the accuracy of the parameter estimates of active users limit the overall system capacity considerably. The model adopted is that of a random access CDMA system where a *single access signature* is available. Note that although all users use the same signature, the system has multiaccess capability due to the asynchronous transmissions of different users. The problem of efficiently detecting the presence of accessing users and acquiring their parameters is discussed. The system capacity, which is defined as the maximum average number of users that can successfully establish reliable connections with the system during a communication period, is found to be quite limited and in particular less than that of the sum capacity of parallel slotted ALOHA channels with the same total bandwidth of the CDMA system.

II System Description

We consider a CDMA system where multiple users can attempt to access the system at the same time. The communication format is to have an access period followed by the information transfer of the users that successfully access the system. The users enter the system, transmit their messages and leave before the system announces the next access start time, thus no other connection is present or established during the service of these users. We assume that all potential users have acquired the base station's pilot signal and are tuned to a downlink paging channel where they can receive broadcast messages. The start time of the access message is broadcasted from the base station along with other necessary access parameters. The delay uncertainty of the new users thus comes from their transmission (propagation) delays relative to the broadcast of the base station. We assume these delays to be

less than 1 bit period for each new user (see Figure 1).

The base station has to detect the activity of *a random number* of users along with the delays of each of these users during the access period. In [8], this is called the "multiuser access detection" problem. While the major difference between multiuser access detection and the multiuser timing acquisition is the uncertainty about the activity status' of all the users, one should also note the stringent requirement on the time frame in which the access has to be completed. In particular, the acquisition time should be much less than the duration the information bits which is likely be on the order of a few of hundreds of bits. This requirement precludes the use of extensions of some recently proposed algorithms [6, 9, 10] to the case where the number of users is unknown.

We assume the initial packet each user sends includes a preamble (a sequence of 1s) that will be used to detect the user's activity and estimate its arrival time followed by the user's identification. If the user's presence is *detected* by the system during the access phase and if the user can establish a reliable connection, the user receives an acknowledgment to go forward with the information transmission. Note that, alternatively, one can think of a system where the preamble is immediately followed by the information bits and the user receives an acknowledgment if the information is received correctly. The capacity of both systems is the same if the requirement to have a successful packet transmission is to have the same quality of service as the identification sequence in the previous system.

The described model suggests a *two-stage* receiver whose initial stage works on the transmitted preambles to detect the activity status of the users and is followed by a detector which will decode the active users' identification information using the findings of the first stage. The performance of the first stage is of vital importance to the system since the performance of the second stage detector hinges upon the correctness of the information supplied by the first stage. A false alarm event, the event that the system erroneously declares a user present when there is none, implies a waste of resources for the second stage since it may require the detector to try to decode fictitious users and to suppress their actually non-existing interference to other users. A miss event, the event that the system fails to capture a user, is also highly undesirable since an active user will not enter the system and its interference to the other users will not be cancelled during the second stage.

III Multiuser Access Detection

Let us first concentrate on the first stage of the detection process, that is designing the Multiuser Access Detector (MUAD). Since the first access stage uses a preamble of all 1's, the received signal during the first stage of access is

$$r(t) = \sum_{i=1}^{N_A} \sqrt{q_i} s_a(t - \tau_i) + n(t) \qquad t \in [0, LT_b] \qquad (1)$$

where N_A is the number of active users, q_i and τ_i are the received power and the delay of the user i, and $n(t)$ is the zero mean white Gaussian noise with power spectral density σ^2, $L > 1$ is the length of the preamble in bits, and T_b is the bit duration. We assume $0 \leq \tau_i \leq T_{max} < T_b$. The accessing signature sequence $s_a(t)$ can be expressed as

$$s_a(t) = \sum_{i=0}^{G-1} c(i) \frac{1}{\sqrt{G}} p(t - iT_c) \qquad (2)$$

where G is the processing gain, T_c is the chip duration, $c(i) \in \{-1, 1\}$ is the ith chip value, and $p(t)$ is the chip waveform normalized to have unit energy. Throughout the paper we will assume for simplicity that $p(t)$ is rectangular. The received signal is observed from the start of the access message with 1-bit delay, thus for a total of $L - 1$ bits. Since $0 \leq \tau_i \leq T_{max} < T_b$, observing the signal with 1-bit delay ensures the capture of at least one bit period where all new users are actively sending their access preamble. Note that during each observed bit interval, the contribution of each active terminal consists of the access signature sequence circularly shifted by that terminal's delay value (see Figure 2).

For further processing, the received signal is discretized by projecting onto a sequence of chip waveforms delayed by multiples of T_c and then sampling at the chip rate (T_c samples per second). The idea then is to separate users (by processing the discrete signal) whose delays are sufficiently apart from each other. At this point, one can envision sampling the filter output every δ second (a fraction of T_c) in the hopes of resolving users that are closer in their delay values. For rectangular chip waveforms, sampling at a higher rate effectively would increase the dimension of the signal space and would improve resolvability. However, it is observed in [11] that this higher dimensional space is an artifact resulting from the use of rectangular pulses. In practice, when chip pulses are bandlimited, there is no such improvement in resolvability. Since our reasoning for using rectangular pulses is strictly for analytical ease, we will purposefully refrain from this faster processing.

Let us move on to the discrete signal representation. It is easy to see that with chip waveforms time limited to duration T_c, one can express a chip matched filtered signal delayed by any delay value as a combination of two *adjacent* vectors as explained below.

Let us define $s_j^\alpha(t)$ as the circularly shifted version of the basic access signature sequence by $(j + \alpha) T_c$ where $\alpha \in [0, 1]$ is the delay mismatch expressed as a fraction of one chip. The chip matched filtered version of $s_j^\alpha(t)$, $\mathbf{s}_j^\alpha$ can be written as

$$\mathbf{s}_j^\alpha = R(\alpha) \, \mathbf{s}_j + R(1 - \alpha) \, \mathbf{s}_{j+1} = (1 - \alpha) \, \mathbf{s}_j + \alpha \, \mathbf{s}_{j+1} \qquad (3)$$

where $R(\alpha) = 1 - |\alpha|$ is the autocorrelation function of the rectangular chip waveform and $\mathbf{s}_j$ is the access signature sequence circularly shifted by j chips, i.e. $\mathbf{s}_0 = [c(0), \cdots, c(G-1)]$ and $\mathbf{s}_j = [c(G-j), \cdots, c(G-1), c(0), c(1), \cdots, (G-1-j)]$ for $1 \leq j < G$. Associated with each active user k, there is

a fractional mismatch α_k ($0 \le \alpha_k < 1$) such that

$$\tau_k = (i_k + \alpha_k)T_c \qquad i_k = \lfloor \tau_k / T_c \rfloor \qquad (4)$$

Recall that the received signal is observed starting at time $t = T_b$ and the contribution of user k within each bit of the observed signal is the access signature sequence circularly shifted by τ_k. Let us assume $\lceil T_{max}/T_c \rceil = M - 1$ where $M \le G$ and define B_i as the set of users whose delays (τ_k) are in the interval $[(i-1)T_c, iT_c]$ ($i = 1, \cdots, M-1$). The received signal in the observation window $t \in [T_b, L\,T_b]$ then can be expressed as

$$r(t) = \sum_{i=1}^{M-1} \sum_{k \in B_i} \sqrt{q_k} s_{i-1}^{\alpha_k}(t) + n(t) \qquad (5)$$

Now, using (3), we can express the discretized received signal in the jth observation interval ($[j\,T_b, (j+1)\,T_b]$) as

$$\mathbf{r}_j = \mathbf{S v} + \mathbf{n}_j \qquad (6)$$

where $\mathbf{n}_j$ is a white Gaussian vector, $\mathbf{S}$ is the matrix with columns $\mathbf{s}_0, ..., \mathbf{s}_{M-1}$ and

$$v_i = \begin{cases} \sum_{k \in B_i} (1 - \alpha_k)\sqrt{q_k}, & i = 1 \\ \sum_{k \in B_{i-1}} \alpha_k \sqrt{q_k} + \sum_{k \in B_i} (1 - \alpha_k)\sqrt{q_k}, & 2 \le i < M \\ \sum_{k \in B_{i-1}} \alpha_k \sqrt{q_k}, & i = M \end{cases} \qquad (7)$$

As can be seen from (7), users that belong to $B_{(i-1)}$ and B_i contribute to v_i with associated weights that depend on their delay mismatch parameters. Note that it is easy to extend this model to $M = G + 1$ ($\lceil T_{max}/T_c \rceil = G$) by taking into account the fact that in addition to what is given in (7), v_1 and v_M will be one and the same and will have contributions from users in B_{M-1} and B_1. Fixing the maximum number of users to N at each interval, we can express $\mathbf{v}$ as

$$\mathbf{v} = \mathbf{A D b} \qquad (8)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 - \alpha_1 & 1 - \alpha_2 & \cdots & 0 & \cdots & 0 \\ \alpha_1 & \alpha_2 & \cdots & 1 - \alpha_{N+1} & \cdots & 0 \\ 0 & 0 & \cdots & \alpha_{N+1} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 - \alpha_{NM} \\ 0 & 0 & \cdots & 0 & \cdots & \alpha_{NM} \end{bmatrix} \qquad (9)$$

is the $M \times NM$ matrix that contains the fractional mismatch values of all possible users (active or not), $\mathbf{D}$ is the $NM \times NM$ diagonal matrix with the amplitudes as the diagonal entries and $\mathbf{b}$ is the binary ($0 - 1$) NM-vector whose first N components represent the activity status of the possible users that belong to B_1, the next N components represent the users in B_2

and so on. In particular, $b_{(i-1)N+j} = 1$ means that the jth user in B_i is active. Collecting multiple observations and defining $\mathbf{r} = \frac{1}{L-1} \sum_{j=1}^{L} \mathbf{r}_j$, the joint maximum likelihood detector for recognition of active users and estimation of their amplitudes and delays is given by

$$[\hat{\mathbf{A}}, \hat{\mathbf{D}}, \hat{\mathbf{b}}] = \arg \min_{\mathbf{A}, \mathbf{D}, \mathbf{b}} \mathbf{b}^T \mathbf{D}^T \mathbf{A}^T \mathbf{S}^T \mathbf{S A D b} - 2\mathbf{r}^T \mathbf{S A D b} \qquad (10)$$

It is not hard to see that if $\mathbf{A}$ and $\mathbf{D}$ were exactly known, to estimate $\mathbf{b}$ would require to solve a $0 - 1$ quadratic program which is N-P hard. When estimation of $\mathbf{A}$ is also required, even with perfect knowledge of the amplitude levels, the mismatch values do not have a unique solution if there is more than one active user in a given interval. Let us concentrate on the case of equal (known) received power, i.e. $\mathbf{D} = \sqrt{q}\mathbf{I}$ (A similar argument can be made for the unequal power case). Due to the discrete nature of $\mathbf{b}$, the solution to (10) has to be found by estimating $\mathbf{A}$ for every distinct possibility of users' activities represented by the $\mathbf{b}$ vectors. First, we observe that order of users in an interval is not important. For instance, if b_i and b_{i+1} are exchanged the same cost value is attained by exchanging the corresponding columns of $\mathbf{A}$, i.e. $\mathbf{a}_i$ and $\mathbf{a}_{i+1}$. Secondly, given $\mathbf{b}$, the minimizer of (10) can be found in the variable

$$\tilde{\mathbf{v}} = \mathbf{A b} = \sum_{i=1}^{NM} b_i \mathbf{a}_i = \sum_{i, b_i = 1} \mathbf{a}_i$$

and not in individual $\mathbf{a}_i$'s. Now recall the structure of $\mathbf{A}$ in (9). Since only the sum of the mismatch values of the users show up in $\mathbf{v}$, if there is more than one column that corresponds to the same interval, it is not possible to uniquely identify the individual mismatch values. The conclusion here is that by using the discrete delay filters we can have a maximum likelihood estimate of how many active users arrive at each interval and the sum of the mismatch values of these users. This has to be done by solving $(N+1)^M$ optimization problems. Furthermore, if the number of users arriving at each interval is more than one, then we have effectively a collision since their individual delays can not be identified.

Since the above optimization is prohibitively complex for large M, one should consider suboptimum methods. A decorrelating multiuser access detector was considered in [8]. The detector was derived using circular matched filters, but the formulation is equivalent to the model here. The idea is to eliminate users that are outside $[(i-2)T_c, iT_c]$ at each detector output y_i. The output of the detector is

$$\mathbf{y} = \frac{1}{(L-1)} (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S r} = \mathbf{v} + \tilde{\mathbf{n}} \qquad (11)$$

where $\tilde{\mathbf{n}}$ is 0-mean Gaussian noise vector with covariance $\frac{\sigma^2}{(L-1)} \Gamma^{-1}$ where we denote $\Gamma = \mathbf{S}^T \mathbf{S}$. Since in general, v_i contains contributions from an unknown number of users each of which has an unknown delay offset, it is difficult to solve for the exact number of active users. Instead, we can simply try to

detect if there are *any* active users around the ith delay value by designing a simple threshold test. If we ignore the correlations between filter outputs, we can design a Neyman-Pearson type binary hypothesis test for y_i with the hypotheses:

- H_{i1} : Activity detected at the ith filter
- H_{i0} : No activity at the ith filter

The threshold is set for each test so that a fixed false alarm probability performance (α_F) is achieved. Specifically, for y_i we set the threshold Λ_i such that

$$\alpha_F = \int_{\Lambda_i}^{\infty} \frac{1}{\sqrt{2\pi\Gamma_{ii}^{-1}\sigma^2/(L-1)}} e^{-x^2/(2\Gamma_{ii}^{-1}\sigma^2/(L-1))} dx \quad (12)$$

The resulting threshold Λ_i and the test are

$$\Lambda_i = Q^{-1}(\alpha_F)\sqrt{\frac{\Gamma_{ii}^{-1}\sigma^2}{(L-1)}} \qquad y_i \underset{H_{0i}}{\overset{H_{1i}}{\gtrless}} \Lambda_i \quad (13)$$

where $Q(x)$ is the standard normal complementary CDF.

So, for activity to be detected at the output of the ith decorrelator, the output should exceed a factor that depends on the specified false alarm rate times the *enhanced* noise magnitude at that decorrelator. If activity is detected at the output of the ith decorrelator, the second stage detector should use this information to try and decode the identification information of a user around the corresponding delay value. The second stage detector, as explained before is a multiuser bit detector.

IV System Capacity

The capacity is defined as the average number of users that can successfully access the system during an access attempt. This depends on the performance of the MUAD and the performance of the second stage bit detector which clearly is coupled with that of the MUAD under given system parameters which include the user arrival rate, MUAD false alarm rate and the quality of service requirement for reliable connection to the system. Since the exact analytical expression for the success probability of a user is complex, we have simulated a DS/CDMA system with processing gain $G = 31$ to observe the average number of users that the system can accommodate. The accessing signature sequence is chosen to be a basic M-sequence. The system has Poisson arrivals with mean λ and users which cannot access the system are immediately cleared. We assume equal received power for all users with values $10, 13, 15$ and 20dB and the background noise power σ^2 to be unity (0dB). In all experiments, the thresholds for the decorrelator outputs of the first stage are set such that the false alarm rate of each test is $\alpha_F = 0.01$. The second stage receivers constructed are one-shot asynchronous decorrelators. For all experiments, the reliable connection requirement is that the signal-to-interference ratio (SIR) of the user has to be above 5dB.

The system we compare against is a slotted ALOHA system with M parallel orthogonal (narrowband) channels where

$T_{max}/T_c = M - 1$. Assuming the channels are in frequency domain, we have an equivalent system that occupy the same bandwidth that of the CDMA system if we set $M = G$. It is easy to see that with the given received power values and SIR target, the ALOHA channels are only limited by collisions. We again assume unsuccessful access attempts are immediately cleared. Recall that we assume uniform delay values for the CDMA users and thus the probability of a user being in any one chip interval is $1/M$. In the equivalent ALOHA model, we assume the probability of a user transmitting in a particular channel is $1/M$. This enables us to decompose the aggregate arrival rate into M independent λ/M Poisson streams each of which is the traffic load of one ALOHA channel. Thus, the sum capacity is M times that of one ALOHA channel. The total throughput (average number of users that can access the system) is $\lambda e^{-\lambda/M}$. The maximum is M/e.

Our experiments showed that in the absence of a fine tracking algorithm before the second stage, the system capacity is very low. This is due to the sensitivity of the bit detectors used in the second stage to the difference between the exact delays of the active users and the coarse delay estimates supplied by the first stage. Furthermore, even in the presence of tracking, where we assumed users can be tracked perfectly when alone in a chip interval, we observed the capacity is lower than sum capacity of M parallel slotted ALOHA channels (Figures (3) and (4)). This effect is even more pronounced when the delay space covers the whole bit interval where the equivalent ALOHA system with 31 orthogonal channels can support about one more user than the packet CDMA system even with very high received power values (Figure (4)).

There may be a number of ways to increase the capacity of the CDMA system. Higher layer solution possibilities include developing a frame structure with more than one access period before data transmission, designing retry strategies that will separate the users further apart in delay space and/or distribute the traffic load between multiple access attempts. For example, since the mobile already has the delay estimate for the downlink (through base station pilot acquisition), it can use this information to change its delay during retransmissions. Nevertheless, our observations lead us to believe that the access process is capacity limiting for a packet switching CDMA system. This is in contrast to earlier literature where capacity of random access CDMA is characterized essentially ignoring the effects of user acquisition.

References

[1] F. Adachi, M. Sawahashi, and H. Suda. Wideband DS-CDMA for next-generation mobile communication systems. *IEEE Comm. Mag.*, pages 56–69, September 1998.

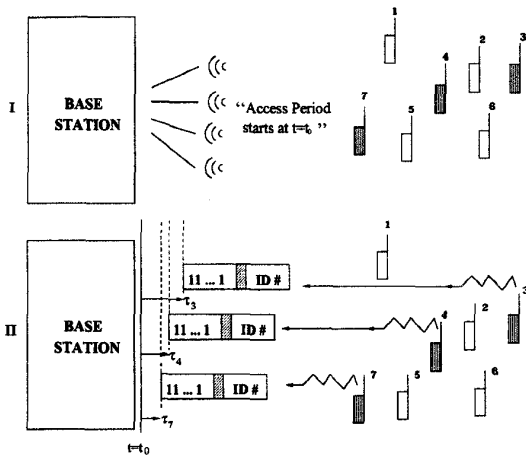[2] Code division multiple access networks - IV. *IEEE JSAC*, 14(9), December 1996.

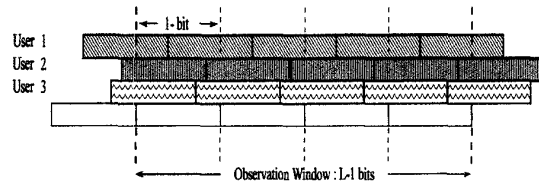Figure 1: Multiaccess Model for the packet switched CDMA system

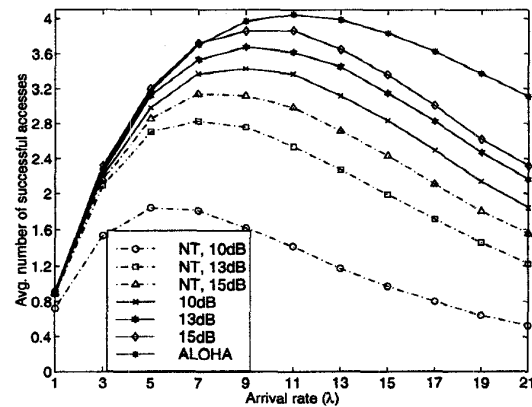Figure 2: Observation of the access preamble by the multiuser access receiver

Figure 3: Comparison of access capacity for received power values 10,13,15dB and M orthogonal ALOHA channels, NT:No tracking, M=11, L=11 (Longer preambles exhibited similar performance)
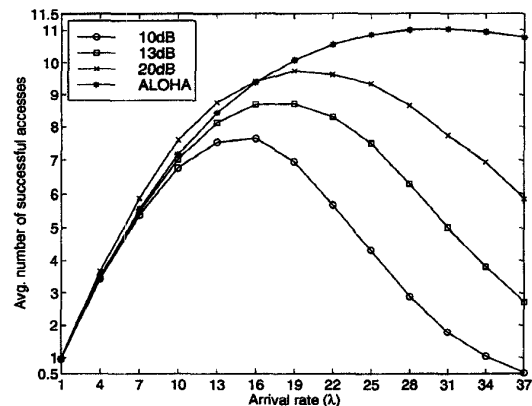

Figure 4: Comparison of access capacity for received power values 10,13,20dB and M orthogonal ALOHA channels, M=G=31, L=11

[3] N. Abramson. Fundamentals of packet multiple access for satellite networks. *IEEE JSAC*, 10(2), 1992.

[4] F.Khan, C.Roobol, and J.Larsson. Performance of a common channel packet access in WCDMA. In *PIMRC 98'*, 1998.

[5] S. Dastangoo, S. Resheff, and B. R. Vojcic. Performance analysis of multi-code spread ALOHA system: Finite and infinite population. In *PIMRC 98'*, 1998.

[6] S. E. Bensley and B. Aazhang. Subspace-based channel estimation for code division multiple access communication systems. *IEEE Trans. Comm.*, 44(8):1009–1020, 1996.

[7] U. Madhow and M. B. Pursley. Acquisition in direct-sequence spread-spectrum communication networks: An asymptotic analysis. *IEEE Trans. Info. Theory*, 39(3):903–912, May 1993.

[8] A. Yener and R.D. Yates. Multiuser access detection for CDMA systems. In *CISS'98*, pages 17–22, 1998.

[9] M. Pallas and G. Jourdain. Active high resolution time delay estimation for large BT signals. *IEEE Trans. Signal Proc.*, 39(4):781–788, April 1991.

[10] E. G. Strom et. al. Propagation delay estimation in asynchronous direct-sequence code-division multiple access systems. *IEEE Trans. on Comm.*, 44(1):84–93, 1996.

[11] A. Mantravadi and V. V. Veeravalli. On discrete sufficient statistics for asynchronous band-limited CDMA systems. *IEEE Trans. Comm.* Submitted December 1998. Available at http://www.ee.cornell.edu/ venu/.