Delta Modulation based Prediction for Access Control in Integrated Voice/Data CDMA Systems

Cristina Comaniciu and Narayan B. Mandayam
Wireless Information Network Laboratory (WINLAB), Rutgers University
73 Brett Road, Piscataway, NJ 08854-8060
EMail: ccris,narayan@winlab.rutgers.edu

Abstract

In this paper, we propose access control protocols for integrated voice/data CDMA systems that are based on estimating the residual capacity available for data users. The data model considered in the analysis is suitable for non-transparent services (e.g. email, file transfer, store-and forward facsimile, etc). The residual capacity is derived from the feasibility condition for power control using an adaptive prediction technique based on delta modulation. The analysis is extended for the case of imperfect power control. Two access control protocols are considered: Modified Delta Modulation with Scheduled access (MDM–S) and Modified Delta Modulation with Random access (MDM–R). Comparisons with earlier proposed protocols reveal a better performance in terms of average throughput for data, for a given system capacity and outage probability.

This paper was presented in part at ICC'99, June 1999, Vancouver, Canada. This work is supported in part by the NSF under Grant No. NCR 97-06036.

1 Introduction

In recent years, there has been tremendous interest in designing access control for wireless multimedia traffic. The need for an access control protocol comes from the complementary QoS requirements that voice and data have: voice requires a continuous bit stream service, is intolerant to delays and can tolerate occasional loss, whereas data is error intolerant, suitable for discontinuous packetized transmission and tolerant to moderate delays. The access protocol maintains the total system interference below a certain level by scheduling the data users to transmit more packets in periods of low voice activity and reducing the transmission when the voice load is high. Earlier work on integrated voice/data CDMA systems includes [3], where data users use a slotted ALOHA protocol with retransmission control to contend for resources that the voice users have not requested. The analyzed access strategy neglects the voice activity factor. In [7], multiple classes of users, having different information rates and different QoS requirements, are considered for a resource allocation optimization in a DS-CDMA system. The QoS is controlled through the selection of the processing gain and of the transmitting power. A class of protocols based on updating the access probability according to a persistence state value, which is increased if a predefined outage condition is violated and decreased otherwise, is analyzed in [6], [8], [10] for different outage conditions and data models. A comparative analysis of control procedures based on persistence state and optimum prediction is presented in [10]. It is shown that prediction schemes provide an upper bound on performance but have a high implementation complexity, whereas the persistence state based procedure gives suboptimal performance results but has the advantage of a simple, practical implementation.

Another approach in designing an access control scheme for integrated voice/data systems is presented in [12]. After admission, the data users are separated into two categories: standby and active. A standby user is allowed to be active with a probability p_{start} computed as a function of the residual capacity. The central station updates the probability at each time slot and transmits the new value to the standby data users. The data users who gain access then proceed to transmit until all their packets are exhausted. This differs from the scheme in [10] where each user transmits in each time slot with some probability that reflects the current system load. The design in [10] and in [12] also differs in the technique used to compute and update the access probability.

In this paper we consider the problem of improving the congestion control procedure such that higher data throughput is obtained for a given system operating point (bidimensional capacity for an outage probability constraint). The analysis distinguishes three causes for outage: imperfections in estimating the residual capacity, imperfections in scheduling the desired number of data users and imperfections in power control. The goal is to design an access control protocol which eliminates the outage due to imperfections in estimating the residual capacity, thus achieving zero outage probability for the perfect power control case with scheduled access. A Modified Delta Modulation (MDM) algorithm is proposed for estimating the residual capacity and two different access protocols are analyzed: MDM–S (Modified Delta Modulation with Scheduled access) and MDM–R (Modified Delta Modulation with Random access). Results are compared to those of the persistence state based access control [10] and it is shown that both MDM–S and MDM–R perform better. The prediction scheme proposed for residual capacity estimation updates the access probability using just the sign of the difference between two consecutive samples. This property, together with the fact that the update frequency is very low, yield a very efficient downlink utilization for MDM–R.

In section 2 we introduce the system model. In section 3 the MDM–S and MDM–R access control protocols are presented for the case of perfect power control. The analysis is extended in section 4 for the case of imperfect power control. An optimal admission control procedure is discussed in section 5 and concluding remarks are presented in section 6.

2 System Model

The access control protocol is proposed for the reverse link of a direct sequence single cell CDMA system with bandwidth W. The uplink is assumed to be slotted. There are two types of users: voice and data, having different information rates R_v and R_d reflected in different spreading gains for the system, G_v, G_d. An ON/OFF voice activity model is assumed, with a transition rate μ from ON to OFF and λ from OFF to ON states. In real scenarios, the users' activity is characterized by multiple rates. Note that the analysis can be extended to incorporate this more complex model. However, to illustrate the protocol performance, the simplified model suffices.

The cumulative voice activity process is approximated as a discrete Markov chain. It is assumed that the duration of a slot d is small compared with $1/\lambda$ and $1/\mu$ such that the probability of two or more events in a time slot is negligible. Considering K_v voice users in the system and $v(n)$ active voice users in the n^{th} time slot, the transition probabilities [10] are given by:

$$P\{v(n+1)=k+1|v(n)=k\}=\frac{\lambda_k}{\lambda_k+\mu_k}(1-\exp\left(-(\lambda_k+\mu_k)d\right))\ , \tag{1}$$

$$P\{v(n+1)=k-1|v(n)=k\}=\frac{\mu_k}{\lambda_k+\mu_k}(1-\exp\left(-(\lambda_k+\mu_k)d\right))\ , \tag{2}$$

$$P\{v(n+1)=k|v(n)=k\}=\exp\left(-(\lambda_k+\mu_k)d\right)\ , \tag{3}$$

where: $\frac{1}{\lambda_k}=\frac{1}{\lambda(K_v-k)}$, $\frac{1}{\mu_k}=\frac{1}{\mu k}$.

Thus, the stationary probability of state k, characterized by k active voice users is:

$$P_V(k)=\frac{\begin{pmatrix} K_v \\ k \end{pmatrix}\left(\frac{\lambda}{\mu}\right)^k}{\sum\limits_{j=0}^{K_v}\begin{pmatrix} K_v \\ j \end{pmatrix}\left(\frac{\lambda}{\mu}\right)^j}. \tag{4}$$

The number K_d of data users in the system is deterministic and each user offers exactly one data packet/slot. This data model is suitable for non-transparent services (e.g. email, file transfer, store-and forward facsimile, etc) [5]. The number of voice and data users in the system is limited to its bidimensional capacity by an admission protocol. This capacity can be determined exactly for a given probability of outage and a specified data access delay. A data packet blocked by the access protocol, or errored, remains in the buffer to be retransmitted. It is assumed that a packet is received correctly whenever the outage condition is satisfied and errored otherwise. The data activity is a consequence of the access protocol. A data user is active in the current time slot, if it is allowed to transmit.

The access protocol design is based on estimating the system residual capacity using a modified delta modulation algorithm. The residual capacity is determined according to an outage condition, defined such as to reflect the error probability requirements for both voice and data users. The performance measures considered are: Outage Probability P_{out} (the fraction of time the outage condition is violated), Throughput T_d (mean number of transmitted packets/slot per data user), Goodput G_d (mean number of correctly received packets/slot per data user), Mean Access Delay D_a (average delay, in slots, experienced by a data packet until is allowed to transmit), Mean Transmission Delay D_t (average delay, in slots, experienced by a data packet until is correctly received), Probability of Blocking P_b (average probability of blocking for an arbitrary user) and system capacity.

3 Access Control based on Power Control Feasibility

3.1 Residual Capacity Estimation

In a CDMA system for which the target SIRs of the N users are γ_j and the users' information rates are r_j, it was proved in [4] that the power control feasibility is given by:

$$\sum_{j=1}^{N} \frac{1}{\frac{W}{r_j \gamma_j} + 1} < 1. \tag{5}$$

In a system with perfect power control, the violation of (5) constitutes an outage.

Defining $S(n)$ as the load for $v(n)$ active voice users (having target SIRs γ_v) and $d(n)$ active data users (having target SIRs γ_d) in the n^{th} time slot, and using the notations $a_v = \frac{W}{R_v \gamma_v} + 1$ and $a_d = \frac{W}{R_d \gamma_d} + 1$ [10] the condition (5) becomes:

$$S(n) = \frac{v(n)}{a_v} + \frac{d(n)}{a_d} < 1. \tag{6}$$

By choosing δ to be a very small value (e.g. $\delta = 0.001$), the above condition is guaranteed to be true if:

$$S(n) = \frac{v(n)}{a_v} + \frac{d(n)}{a_d} \leq 1 - \delta. \tag{7}$$

Rearranging (7), we obtain:

$$d(n) \leq a_d(1 - \delta) - \frac{a_d}{a_v} v(n) \;, \tag{8}$$

where $d(n)$ represents the ideal residual capacity of the system, defined as the remaining capacity after the voice contribution is subtracted. In an integrated voice/data CDMA system with perfect power control, violation of (8) constitutes an outage.

Since $v(n)$ is a random variable, only an estimate $\widehat{d(n)}$, as a function of $v(n-1)$ can be derived. The access control protocol attempts to schedule exactly $\lfloor \widehat{d(n)} \rfloor$ users at each time slot. To maintain QoS, the fraction of the time the outage condition is violated should be very small, typically 1%. We distinguish three main causes for outage:

I Imperfections in estimating the residual capacity,

II Imperfections in scheduling the desired number of data users,

III Imperfections in power control.

The goal is to design an access control protocol which eliminates the first listed cause for outage, thus achieving zero outage probability for the perfect power control case with scheduled access. The analysis is extended to consider II (in subsections 3.3 and 4.3) and III in section 4. To determine the residual capacity estimate $\widehat{d(n)}$, a simple, practical, prediction scheme, based on a modified delta modulation algorithm, is proposed. Delta modulation [13] is a technique for converting analog signals $s(t)$ into digital ones $\widehat{s(t)}$ (Figure 1 (a)). The modulated signal is a staircase representation of the analog signal. The accuracy of the approximation is given by two parameters: the sampling frequency, f_s, and the step of the modulation, Δ. There are two types of modulation distortions: (I) granular noise (too coarse quantizer for the flat regions) and (II) slope overload distortion (if the signal $s(t)$ changes too fast, $\widehat{s(t)}$ cannot follow it). The condition for no slope overload distortion is:

$$\left| \frac{ds(t)}{dt} \right| < \Delta f_s. \tag{9}$$

In designing the access control protocol, we use the prediction method of delta modulation to approximate the ideal residual capacity function $d(n)$. $\widehat{d(n)}$ is the staircase approximation of $d(n)$ and the accuracy of the representation is given by f_s (the inverse of the slot duration) and Δ (the step for the staircase approximation function $\widehat{d(n)}$). Specific values for d and Δ for the perfect power control case will be derived later in this section. For the moment, for a general illustration, consider that the values for d and Δ have been chosen such that condition (9) holds.

Since a performance measure for the access control protocol is the outage probability, several modifications for delta modulation are necessary in order to guarantee $\widehat{d(n)} \leq d(n)$. Firstly, a guard margin equal with the step of modulation (Δ) is imposed, such that the function to be approximated becomes $d_t(n) = d(n) - \Delta$, where $d(n)$ is computed as the maximum value for which (8) holds. Moreover, the steps of the DM algorithm are modified as follows:

1. For the first time slot, no access procedure is used: at the end of the first time slot, the number of active voice users is measured and the value $\widehat{d(1)}$ is initialized as $\widehat{d(1)} = d(1) - \Delta$.

2. For each time slot $n > 1$:

- measure the number of active voice users in the previous time slot ($v(n-1)$) and compute $d(n-1) - \Delta$;
- if $\widehat{d(n-1)} > d(n-1) - \Delta$,
  $\widehat{d(n)} = \widehat{d(n-1)} - \Delta$.
- if $\widehat{d(n-1)} < d(n-1) - \Delta$,
  $\widehat{d(n)} = \widehat{d(n-1)} + \Delta$.
- if $\widehat{d(n-1)} = d(n-1) - \Delta$,
  $\widehat{d(n)} = \widehat{d(n-1)}$.
- Employ an access procedure to allow $\lfloor \widehat{d(n)} \rfloor$ users to transmit.

In Figure 1 (b) simulations results are plotted for the more general case of imperfect power control which will be analyzed in section 4. The upper graph illustrates the relationship between the ideal residual capacity and the estimated one. It can be seen that the estimated function never exceeds the ideal residual capacity. The lower graph in 1 (b) illustrates the prediction technique: $d_t(n)$ is the original function and $\widehat{d(n)}$ is the approximation. It can be seen that, on average,

$$\widehat{d(n)} \approx d_t(n) = d(n) - \Delta. \tag{10}$$

An exact relationship can be defined by observing the interconnection among $\widehat{d(n)}$, $d(n)$ and the voice activity model:

$$\widehat{d(n)} = \begin{cases} d(n); & \text{with } p = P(v(n+1) = k+1/v(n) = k) \\ d(n) - \Delta; & \text{with } p = P(v(n+1) = k/v(n) = k) \\ d(n) - 2\Delta; & \text{with } p = P(v(n+1) = k-1/v(n) = k) \end{cases} \tag{11}$$

However, experimental results indicate a good accuracy for (10) due to the fact that $P(v(n+1) = k+1/v(n) = k)$ is approximately equal to $P(v(n+1) = k-1/v(n) = k)$. Therefore, in the remainder of the paper, we will use the approximation from (10).

Since the cumulative voice activity process is modeled as a discrete Markov chain, the slot duration d must be chosen to be sufficiently small such that at most ± 1 variation in the number of active voice users can occur from one time slot to another. As a consequence, having $d(n)$ defined as the maximum value which satisfies (8), $d(n)$ can vary from one slot to another by at most $\pm \frac{a_d}{a_v}$. Therefore, the constraint from (9) gives $\Delta = \frac{a_d}{a_v}$ for the perfect power control case analyzed in this section.

The lower graph in Figure 1 (b) illustrates that the MDM performs identically as the simplest first order linear predictor, i.e.,

$$\widehat{d(n)} = d(n-1) \ , \tag{12}$$

which was proposed for estimating the residual capacity in [12]. However, MDM is more efficient in terms of feedback information from the base station: to update the value of the residual capacity only the sign of the difference between successive samples has to be coded as opposed with coding and transmitting an entire sample. The downlink utilization efficiency for MDM–R is discussed in section 3.3. Moreover, without imposing a guard margin (Δ) an outage occurs each time the number of active voice users increases, such that the minimum achievable outage probability (without considering imperfections in the power control or scheduling imperfections due to the access method) using (12) is:

$$P_{out} = \sum_{k=0}^{K_v} \left(P\{v(n+1) = k+1 | v(n) = k\} \right) P_V(k) \ , \tag{13}$$

e.g. for $K_v = 10$ and voice activity 0.4 the outage probability due to imperfections in estimating the residual capacity is 7.4%. The proposed prediction scheme using MDM with a tolerance guard margin guarantees zero outage probability when only type I outage source is considered in the analysis.

3.2 Performance of MDM–S

For MDM–S, the base station will maintain a round fair list with all data users admitted in the system and will signal users to transmit or not, using an 1 bit access flag: "1" if the user is allowed to transmit in the next time slot, "0" if not. A drawback for this procedure is that the access bit cannot be broadcast, a possible solution will be for the users to listen to a dedicated fraction of a time slot to extract the proper access bit.

The proposed algorithm guarantees $\widehat{d(n)} \leq d(n)$, thus the access control protocol will never schedule more users than the available residual capacity $d(n)$. In the case of perfect power control, the access procedure causes no outage, i.e., MDM–S gives zero outage probability.

The mean number of data users scheduled to transmit in a time slot is given by the average residual capacity:

$$E[\lfloor \widehat{d(n)} \rfloor] \approx E[\lfloor (a_d(1-\delta) - \frac{a_d}{a_v} v(n) - \Delta) \rfloor] = \sum_{v(n)=0}^{K_v} \lfloor (a_d(1-\delta) - \frac{a_d}{a_v} v(n) - \Delta) \rfloor P_V(v(n)). \tag{14}$$

Assuming a fair division of time, on average, each data user will have a mean "duty cycle" of $\frac{E[\lfloor \widehat{d(n)} \rfloor]}{K_d}$, thus the throughput in packets/slot is:

$$T_d = \begin{cases} \frac{1}{K_d} \sum_{v(n)=0}^{K_v} \lfloor (a_d(1-\delta) - \frac{a_d}{a_v} v(n) - \Delta) \rfloor P_V(v(n)), & \text{if } E[\lfloor \widehat{d(n)} \rfloor] \leq K_d \\ 1, & \text{if } E[\lfloor \widehat{d(n)} \rfloor] > K_d \end{cases} . \quad (15)$$

The Mean Access Delay for data can be computed by considering that on average, each data user waits $\frac{1}{T_d} - 1$ slots until it is allowed to transmit a packet, therefore:

$$D_a = \frac{K_d}{E[\lfloor \widehat{d(n)} \rfloor]} - 1. \quad (16)$$

Using the assumption that the error and the outage are correlated (see section 2.), the mean access delay and mean transmission delay are identical, due to the zero outage probability property. This also implies equality between throughput and goodput.

In [11] a scheme based on scheduling was proposed for maximizing the data users' throughput. As opposed to [10] and [12], the work in [11] does not attempt to develop an access control for balancing the system load according to the voice activity fluctuations. The analysis in [11] proves that the reduced interference seen by the users as a consequence of scheduling, allows transmission at higher bit rates.

The approach here is different. Data users are scheduled to efficiently use the available residual resources. The number of data users allowed to transmit in each time slot is restricted by the number of active voice users such that the interference level is bounded. Since each data user is transmitting at most one packet per slot, the throughput will decrease with the increase in system capacity. The system QoS specifications for the data users' throughput are: minimum rate R_{min}, average rate R_{av} and maximum rate R_{max}.

In order to guarantee a minimum throughput rate the condition $\lfloor a_d(1-\delta) - \frac{a_d}{a_v} K_v - \Delta \rfloor > 0$ should hold (resources are available for data even if all voice users are active). As a consequence with a data transmission rate of $R_d = \frac{W}{G_d}$ the minimum throughput rate is:

$$R_{min} = \frac{W}{G_d K_d} \lfloor (a_d(1-\delta) - \frac{a_d}{a_v} K_v - \Delta) \rfloor. \quad (17)$$

The maximum throughput rate is obtained for $T_d = 1$:

$$R_{max} = R_d = \frac{W}{G_d} , \quad (18)$$

and the average throughput rate is given by:

$$R_{av} = R_d T_d = \frac{W}{G_d} T_d. \tag{19}$$

Numerical results are presented in section 3.4.

3.3 Performance of MDM–R

An alternative to the scheduling procedure is to assign at each time slot n an access probability $p(n)$ to all data users. We determine $p(n)$ as a maximum likelihood estimate such that it maximizes the probability that exactly $\lfloor \widehat{d(n)} \rfloor$ data users will transmit in the next time slot:

$$P\{\lfloor \widehat{d(n)} \rfloor | p(n)\} = \begin{pmatrix} K_d \\ \lfloor \widehat{d(n)} \rfloor \end{pmatrix} p(n)^{(\lfloor \widehat{d(n)} \rfloor)} (1 - p(n))^{(K_d - \lfloor \widehat{d(n)} \rfloor)} \tag{20}$$

$$\frac{\partial P\{\lfloor \widehat{d(n)} \rfloor | p(n)\}}{\partial p(n)} = 0 \Rightarrow p(n) = \frac{\lfloor \widehat{d(n)} \rfloor}{K_d}. \tag{21}$$

$p(n)$ will be updated at each time slot, by at most $\frac{\pm\Delta}{K_d}$, therefore the system will need to broadcast a 2 bit feedback, representing $\{+1\}$,$\{-1\}$ or $\{0\}$. An efficient way of coding the feedback information arises from two observations:

- Only 2 out of 4 possible bits combinations are useful for access probability updates;
- Most of the time, there is no need to update $p(n)$ (see Figure 1 (b)).

Thus, a flag bit signals if the next bit contains update information (± 1), or if the two bits represents something else. The update of $p(n)$ is needed only when the number of active voice users changes from one slot to another. Hence, the probability that an update for $p(n)$ is necessary is given by:

$$P_u = \sum_{k=0}^{K_v} \left[P\{v(n+1) = k+1 | v(n) = k\} + P\{v(n+1) = k-1 | v(n) = k\} \right] P_V(k), \tag{22}$$

$$P_u = \sum_{k=0}^{K_v} \left[1 - \exp\left(-(\lambda_k + \mu_k) d)\right) \right] P_V(k).$$

Therefore, the average number of bits used for probability access updates is:

$$B_u = 2P_u. \tag{23}$$

The outage probability is the probability that more than $\lfloor \widehat{d(n)} \rfloor$ data users will win the competition and will transmit in the next time slot, averaged over all possible states of the cumulative voice activity discrete Markov chain (all possible values for $v(n)$):

$$P_{out} = E_{v(n)}\left[P\{j > \lfloor \widehat{d(n)} \rfloor\}\right] = \sum_{v(n)=0}^{K_v} \sum_{j=\lfloor \widehat{d(n)} \rfloor+1}^{K_d} \binom{K_d}{j} p(n)^j (1-p(n))^{(K_d-j)} P_V(v(n)) \ , \tag{24}$$

where $p(n)$ is given by (21) and the dependence on $v(n)$ is through $\widehat{d(n)}$, given by (10) and (8). The average blocking probability is given by:

$$P_b = \sum_{v(n)=0}^{K_v} (1-p(v(n)))P_V(v(n)). \tag{25}$$

Each data user will try to transmit at each time slot with probability $1-P_b$ until it succeeds. The number of trials is a geometric random variable with mean $\frac{1}{1-P_b}$ and thus the mean access delay is:

$$D_a = \frac{1}{1-P_b} - 1 = \frac{P_b}{1-P_b}. \tag{26}$$

Assuming $P_{error} = P_{out}$, let P_t be the probability that either a packet is blocked, or is errored, $P_t = P_b + P_{out} - P_b P_{out}$. Then the mean transmission delay (including retransmissions) is given as:

$$D_t = \frac{P_t}{1-P_t}. \tag{27}$$

The data users' throughput is the mean number of packets/slot transmitted:

$$T_d = \frac{1}{K_d} \sum_{v(n)=0}^{K_v} \sum_{j=0}^{K_d} \left(j \binom{K_d}{j} p(v(n))^j \left(1-p(v(n))\right)^{K_d-j} P_V(v(n)) \right). \tag{28}$$

The goodput of data users is the mean number of packets/slot transmitted in time slots with no outage:

$$G_d = \frac{1}{K_d} \sum_{v(n)=0}^{K_v} \sum_{j=0}^{min(\lfloor \widehat{d(n)} \rfloor, K_d)} \left(j \binom{K_d}{j} p(v(n))^j \left(1-p(v(n))\right)^{K_d-j} P_V(v(n)) \right). \tag{29}$$

The minimum throughput rate is now equal to:

$$R_{min} = \frac{W}{G_d K_d} \sum_{j=0}^{K_d} \left(j \binom{K_d}{j} p(K_v)^j \left(1-p(K_v)\right)^{K_d-j} \right) \tag{30}$$

R_{max} and R_{av} are given by (18) and (19) with T_d computed as in (28).

The access control protocol performance can be fine tuned by introducing a probability tolerance (t_p) parameter. Hence, the general computation formula for $p(n)$ becomes:

$$p(n) = \frac{\lfloor \widehat{d(n)} \rfloor - t_p}{K_d} \tag{31}$$

A positive value of t_p has the effect of decreasing the access probability which results in a smaller probability of outage. Equivalently, for an imposed outage value, decreasing $p(n)$ gives larger capacity for the system. The penalty is an increase in the data access delay.

3.4 Numerical Results and Observations

The numerical values chosen for simulations in this section are: $W = 1.23MHz$, $G_v = 128$, $G_d = 64$, $R_v = 9.6kbps$, $R_d = 19.2kbps$, $\frac{1}{\lambda} = 1.0s$, $\frac{1}{\mu} = 1.5s$, $d = 0.02s$, $\gamma_v = \gamma_d = 7dB$ and $K_v = 10$. The graphs in Figure 2 illustrate analytical results as well as simulations for both MDM–S and MDM-R. The performance results are compared with those of the persistence state based access protocol analyzed in [10]. For the persistence state algorithm only simulation results are presented and the value for the K parameter [10] is chosen to be 5. The performance measures considered are the outage probability, Figure 2 (a), the mean access delay, Figure 2 (b), the average goodput per data user (measured in packets/slot), Figure 2 (c), and the percentage gain in the average goodput rate for an arbitrary data user for the proposed algorithms (MDM–S, MDM–R) compared with the persistence state protocol, Figure 2 (d):

$$R_{gain}(\%) = \frac{R_{av\_MDM-S(R)} - R_{av\_pers}}{R_{av\_pers}} \times 100, \tag{32}$$

As a first observation, we note that the outage probability for MDM–S is zero for all possible values of K_d as discussed in section 3.2, therefore the outage probability for MDM–S is not represented. For MDM–R two values for the probability tolerance parameter (31) are considered: $t_p = 0$ and $t_p = 2$. Analyzing Figure 2 (a) and (b) we can see that increasing the t_p value, decreases the outage probability, with a direct consequence on the mean access delay which increases. However, the mean access delay by itself is not a good performance parameter for data users' goodput. This information should always be correlated with the outage probability since a large outage probability has the effect of dramatically decreasing the goodput, due to retransmissions (also a small outage is necessary to guarantee QoS for the voice users). This fact is illustrated in Figure 2 (c). We note that for the region of interest ($P_{out} \approx 0.01$) one can approximate the mean transmission delay with the mean access delay and the goodput with the throughput.

The parameters $K = 5$ and $t_p = 2$ were chosen such that all three schemes will have a common operating point characterized by $P_{out} \approx 0.01$ ($P_{out} = 0$ for MDM–S) for $K_d = 11$ data users admitted in the system such that the performance can be compared. Choosing larger values for the sensitivity parameters (t_p and K) has the effect of sliding the operating point and therefore increasing the capacity and decreasing the goodput for data.

It can be seen that both MDM–S and MDM–R perform better than the persistence state algorithm. For the region of interest, MDM–R gives more that 5% average throughput gain whereas MDM–S has a gain of almost 30% (Figure 2 (d)). Additionally, MDM–R has a better downlink utilization: the mean number of feedback bits used by MDM–R is $B_u = 0.2956$ bits/slot (23), as opposed to $B_{pers} = 1$ bits/slot needed by the persistence state protocol. The results for the perfect power control case constitute an upper bound on performance, since in reality there may be imperfections in power control loops. The effect of imperfections in the power control is analyzed in the next section.

4 Access Control for Imperfect Power Control

4.1 Residual Capacity Estimation

Due to practical limitations, perfect power control is not feasible for all mobile speeds and propagation conditions. Moreover, there is a dynamic range limitation at the receiver which require a total received power at the base station within 10 dB of the background noise power. Therefore, for a real integrated voice/data CDMA system having received SIRs for voice and data ϵ_i^v, ϵ_i^d, respectively, the outage condition is defined [10] as:

$$Z(n) = \frac{1}{G_v}\sum_{i=0}^{K_v}\nu_i(n)\epsilon_i^v + \frac{1}{G_d}\sum_{j=0}^{K_d}\zeta_j(n)\epsilon_j^d \leq (1-\eta) \ , \tag{33}$$

where $Z(n)$ represents the load of the system for the n^{th} time slot and $\nu_i(n)$, $\zeta_j(n)$ the voice and data activity indicators, i.e., $\nu_i(n)$, $\zeta_j(n) \in \{0,1\}$. We assume that each data user is transmitting long enough such that a combination of open and closed loop power control is possible. Thus a lognormal *pdf* [1] is used to model the distribution of the target SIRs for both voice and data. The design of the access protocol will follow the steps from section 3. To determine $d(n)$ we impose a maximum value ϕ for the conditioned outage probability:

$$P_{out|v,d}(Z(n) > (1-\eta)|v(n), d(n)) \leq \phi. \tag{34}$$

Conditioned on $v(n)$ and $d(n)$, $Z(n)$ is a sum of lognormal random variables and it was shown in [9] that an accurate *pdf* model is a lognormal distribution. To simplify calculation, a gaussian approximation is used at this step (it was shown in [1], that the gaussian approximation yields more optimistic results compared with simulations but the differences remained within 1% error).

Let $A = G_v G_d (1-\eta)$. Then condition (33) can be rewritten as:

$$Z'(n) = G_d v(n)\epsilon_i^v + G_v d(n)\epsilon_i^d < A. \tag{35}$$

Using a normal approximation for Z', (34) is equivalent with:

$$P_{out|v,d}(Z'(n) > A|v(n), d(n)) \leq \phi \Rightarrow Q\left(\frac{A - E[Z'(v(n))]}{\sqrt{var[Z'(v(n))]}}\right) \leq \phi \tag{36}$$

where:

$$E[Z'(v(n))] = G_d v(n) E_{\epsilon_v} + G_v d(n) E_{\epsilon_d} = \xi \ , \tag{37}$$

and

$$var[Z'(v(n))] = G_d^2 v(n)\sigma_{\epsilon_v}^2 + G_v^2 d(n)\sigma_{\epsilon_d}^2 = \theta. \tag{38}$$

Let us denote:

$$\rho = \left[G_v E_{\epsilon_v}\sigma_{\epsilon_d}^2 - G_d E_{\epsilon_d}\sigma_{\epsilon_v}^2\right] \tag{39}$$

Solving the system of equations (37), (38) with the constraint given by (36) and considering (39) the following expression is derived:

$$\theta = \frac{1}{E_{\epsilon_d}}\left(\xi G_v \sigma_{\epsilon_d}^2 - v(n) G_d \rho\right) \leq \frac{(A-\xi)^2}{(Q(\phi)^{-1})^2} \ , \tag{40}$$

The expression can be bounded by choosing $v(n) = 0$, for $\rho > 0$ and $v(n) = K_v$, for $\rho < 0$. In our analysis we assume $\rho > 0$ which is in agreement with the numerical values chosen for simulations. Solving the quadratic equation:

$$\frac{1}{E_{\epsilon_d}}\xi G_v \sigma_{\epsilon_d}^2 = \frac{(A-\xi)^2}{(Q(\phi)^{-1})^2} \ , \tag{41}$$

ξ is determined as:

$$\xi = \frac{AE_{\epsilon_d} + \frac{1}{2}\left(Q(\phi)^{-1}\right)^2 \sigma_{\epsilon_d}^2 G_v - \sqrt{\left(AE_{\epsilon_d} + \frac{1}{2}\left(Q(\phi)^{-1}\right)^2 \sigma_{\epsilon_d}^2 Gv\right)^2 - A^2\left(E_{\epsilon_d}\right)^2}}{E_{\epsilon_d}}. \tag{42}$$

Thus, for $\Delta = \frac{G_d E_{\epsilon_v}}{G_v E_{\epsilon_d}}$, selected according to (9), the time averaged estimated residual capacity can be approximated to be:

$$\widehat{d(n)} \approx \frac{1}{G_v E_{\epsilon_d}}\left[\xi - G_d E_{\epsilon_v} v(n)\right] - \Delta. \tag{43}$$

An exact solution can be obtained for (36) in the particular case: $G_v = G_d = G$, $E_{\epsilon_d} = E_{\epsilon_v} = E_\epsilon$ and $\sigma_{\epsilon_v}^2 = \sigma_{\epsilon_d}^2 = \sigma_\epsilon^2$ with $A = G(1-\eta)$, $E[Z'(v(n))] = v(n)E_\epsilon + d(n)E_\epsilon = \xi_e E_\epsilon$ and $var[Z'(v(n))] = v(n)\sigma_\epsilon^2 + d(n)\sigma_\epsilon^2 = \xi_e \sigma_\epsilon^2$:

$$\xi_e = \frac{AE_\epsilon + \frac{1}{2}\left(Q(\phi)^{-1}\right)^2 \sigma_\epsilon^2 - \sqrt{\left(AE_\epsilon + \frac{1}{2}\left(Q(\phi)^{-1}\right)^2 \sigma_\epsilon^2\right)^2 - A^2\left(E_\epsilon\right)^2}}{(E_\epsilon)^2}. \tag{44}$$

Therefore, the estimated residual capacity is given by:

$$\widehat{d(n)} \approx \xi_e - v(n) - \Delta\ , \tag{45}$$

with $\Delta = 1$ (see condition (9)).

4.2 Performance of MDM–S

The approach in designing the access control protocol is identical with the perfect power control case except that $\widehat{d(n)}$ is given by equation (43). Considering a lognormal distribution for $Z'(n)$ and averaging over all possible values for $v(n)$, the outage probability is:

$$P_{out} = \sum_{v(n)=0}^{K_v} Q\left(\frac{ln(A) - E_{ln}[Z'(v(n))]}{\sqrt{var_{ln}[Z'(v(n))]}}\right) P_V(v(n)) \tag{46}$$

where: $var_{ln}[Z'(v(n))] = ln\left(\frac{var[Z'(v(n))]}{(E[Z'(v(n))])^2} + 1\right)$
and $E_{ln}[Z'(v(n))] = ln(E[Z'(v(n))]) - \frac{1}{2}var_{ln}[Z'(v(n))]$.
The mean number of data users scheduled in a time slot is:

$$E[\lfloor\widehat{d(n)}\rfloor] = \sum_{v(n)=0}^{K_v} \lfloor \frac{1}{G_v E_{\epsilon_d}}(\xi - G_d E_{\epsilon_v} v(n)) - \Delta \rfloor P_V(v(n)). \tag{47}$$

Similar to section 3.2 the throughput is given by:

$$T_d = \begin{cases} \frac{1}{K_d} \sum_{v(n)=0}^{K_v} \lfloor \frac{1}{G_v E_{\epsilon_d}} (\xi - G_d E_{\epsilon_v} v(n)) - \Delta \rfloor P_V(v(n)), & \text{if } E[\lfloor \widehat{d(n)} \rfloor] \leq K_d \\ 1, & \text{if } E[\lfloor \widehat{d(n)} \rfloor] > K_d \end{cases}, \tag{48}$$

and the mean access delay is:

$$D_a = \frac{K_d}{E[\lfloor \widehat{d(n)} \rfloor]} - 1. \tag{49}$$

The minimum throughput rate is now given by:

$$R_{min} = \frac{W}{G_d K_d} \lfloor \frac{1}{G_v E_{\epsilon_d}} (\xi - G_d E_{\epsilon_v} K_v) - \Delta \rfloor. \tag{50}$$

The maximum and average throughput rate are given by the formulas derived in section 3.2 with T_d given by (48).

4.3 Performance of MDM–R

The design procedure is identical with the one described in section 3. $p(n)$ is given by (21), or generally by (31) with $\widehat{d(n)}$ determined using equations (43) and (42) instead of (8) and (10).

For a lognormal model for $Z'(n)$, and averaging over all possible values of $v(n)$ and $d(n)$:

$$P_{out} = \sum_{v(n)=0}^{K_v} \sum_{d(n)=0}^{K_d} Q\left(\frac{ln(A) - E_{ln}[Z'(v(n), d(n))]}{\sqrt{var_{ln}[Z'(v(n), d(n))]}} \right) P_D(d(n)) P_V(v(n)) \ , \tag{51}$$

$$P_D(d(n)) = \left(\begin{array}{c} K_d \\ d(n) \end{array} \right) p(v(n))^{d(n)} (1 - p(v(n)))^{(K_d - d(n))} \ ,$$

where:

$$E[Z'(v(n), d(n))] = v(n) G_d E_{\epsilon_v} + d(n) G_v E_{\epsilon_d} \ ,$$

$$var[Z'(v(n), d(n))] = v(n) G_d^2 \sigma_{\epsilon_v}^2 + d(n) G_v^2 \sigma_{\epsilon_d}^2 \ ,$$

$$var_{ln}[Z'(v(n), d(n))] = ln\left(\frac{var[Z'(v(n), d(n))]}{(E[Z'(v(n), d(n))])^2} + 1 \right) \ ,$$

$$E_{ln}[Z'(v(n), d(n))] = ln(E[Z'(v(n), d(n))]) - \frac{1}{2} var_{ln}[Z'(v(n), d(n))]$$

.

Using the same reasoning as in section 3.3, the blocking probability is defined as:

$$P_b = \sum_{v(n)=0}^{K_v} (1 - p(v(n)))P_V(v(n)) \tag{52}$$

and the mean access delay is therefore given by:

$$D_a = \frac{P_b}{1 - P_b}. \tag{53}$$

Similarly, the mean transmission delay is:

$$D_t = \frac{P_b + P_{out} - P_b P_{out}}{1 - (P_b + P_{out} - P_b P_{out})}. \tag{54}$$

The throughput for an arbitrary data user (T_d) is the mean number of packets transmitted as in (28) with $p(n)$ given by (21) or (31) and (43). The goodput per data user is the number of correctly received packets per data user, averaged over all possible data users' winner patterns and over all possible states of the cumulative voice activity discrete Markov chain. It can be expressed in terms of the throughput T_d as:

$$G_d = T_d(1 - P_{out}). \tag{55}$$

The minimum throughput rate is obtained by considering that all the voice users are active as in (30) with $p(K_v)$ given by (21) or (31) and (43) with $v(n) = K_v$. R_{min} and R_{av} are given by the formulas defined in section 3.2 with the throughput computed for the case considered in this section.

4.4 Numerical Results and Observations

The numerical values chosen for the system parameters in this section are the same as in section 3.4, plus: $E_{\epsilon_v} = E_{\epsilon_d} = 7dB$, $\sigma_{\epsilon_v} = \sigma_{\epsilon_d} = 2.5dB$ and $\phi = 0.02$.

To analyze the proposed schemes one should first select possible operating points for the system (characterized by the number of data and voice users admitted in the system for a constraint of 1% outage probability). It can be seen in Figure 3 (a) that for MDM–S any value for K_d can be an operating point of the system, the final choice is determined by an average data goodput requirement as we will discuss in the next section. Choosing $t_p = 1.5$ for MDM–R yields as possible operating points all $K_d \geq 15$. However, for very large capacity, the data users' goodput is very low. To achieve reasonably high goodput rates for data users, two choices are selected: $K_d = 4$ for MDM–R with $t_p = 0$ and $K_d = 5$ for MDM–R with $t_p = 1$. The performances for both MDM–S and MDM–R are compared with simulations results for the persistence state based access control with $K = 10$. It can be seen that both proposed schemes give higher goodputs for data users for a given operating point. The average data rate goodput gain percentage defined in section 3.4 is illustrated in Figure 3 (d): for $K_d = 4$ approximately 28% for MDM–S and 22% for MDM–R and for $K_d = 5$ approximately 29% for MDM–S and 12% for MDM–R. Detailed discussion on how to choose optimal operating points for given system constraints follows in the next section.

5 Admission Control and System Capacity

The admission control policy consists of accepting a new voice or data call set up request if the new system operating point characterized by the triplet $\{K_v^*, K_d^*, D_a^*\}$ is included in the allowable region of the bidimensional capacity diagram (Figure 4). K_v^* represents the total number of voice users in the system, including the candidate voice call, and K_d^* represents the total number of voice users in the system, including the candidate data call. Although all combinations in this region are possible, there are some optimal operating points which give the most efficient resource utilization (see Figure 4 (c)). The bidimensional system capacity is defined as the maximum number of voice and data users that can be supported for a given average goodput rate for data users (fixed value of data mean transmission delay) while complying with the outage condition 99% of the time.

To determine the bidimensional system capacity formula we use the results derived in previous sections. An estimated system residual capacity was determined in sections 3 and 4 as the maximum value for $d(n)$ in (8) for perfect power control and as (43) for imperfect power control. Imposing an average goodput rate for the data users is equivalent to imposing a data mean transmission delay, ζ. For an outage probability of 1% the goodput and throughput are approximatively equal and so are the mean access and mean transmission delay. For simplicity, the analysis is performed for a constraint on the mean access delay.

For MDM–S, from equations (16) and (49) the number of data users in the system for a given delay can be determined:

$$D_a = \frac{K_d}{E[\lfloor\widehat{d(n)}\rfloor]} - 1 = \zeta \Rightarrow K_d = E[\lfloor\widehat{d(n)}\rfloor](1+\zeta). \tag{56}$$

A similar result can be derived for MDM–R:

$$D_a = \frac{P_b}{1-P_b} = \zeta \Rightarrow P_b = \frac{\zeta}{1+\zeta}. \tag{57}$$

Considering that $P_b = 1 - \frac{E[\lfloor\widehat{d(n)}\rfloor]}{K_d}$, the previous expression becomes:

$$K_d = E[\lfloor\widehat{d(n)}\rfloor](1+\zeta). \tag{58}$$

The functional $K_d = func(K_v)$ is given by (8), (10) and (56) or (58) for perfect power control:

$$K_d = \left(\sum_{v(n)=0}^{K_v} \lfloor(a_d(1-\delta) - \frac{a_d}{a_v}v(n) - \Delta)\rfloor P_V(v(n))\right)(1+\zeta)\ , \tag{59}$$

and by (43) and (56) or (58) for imperfect power control:

$$K_d = \left(\sum_{v(n)=0}^{K_v} \lfloor\frac{1}{G_v E_{\epsilon_d}}[\xi - G_d E_{\epsilon_v} v(n)] - \Delta\rfloor P_V(v(n))\right)(1+\zeta). \tag{60}$$

A surface plot for the bidimensional system capacity for an allowable range of delays is represented in Figure 4 (a) for MDM–S and perfect power control and in Figure 4 (b) for MDM–S and imperfect power control. It can be seen that imperfections in the power control have the effect of reducing the system capacity.

Due to the fact that the number of users admitted in the system is an integer number, the surfaces have a staircase aspect. The optimal operating points for the system are chosen such that K_d^* and K_v^* are maximum whereas the delay D_a^* is minimum. The graph in 4 (c) represents a section through the Figure 4 (b) surface, for a mean access delay $D_a^* = 0.6$.

In practice, the number of voice users admitted in the system is limited by imposing a minimum rate goodput for data users (30)(50).

Similar plots can be obtained for MDM–R. The system capacity will slightly decrease compared to MDM–S because MDM–R includes also outage owed to the access procedure.

6 Conclusion

In this paper, we have proposed access control protocols for integrated voice/data CDMA systems with non-transparent data services such as email, file transfer, store-and forward facsimile, etc. The control strategies estimate the residual capacity available for data users using a prediction technique based on delta modulation. Specifically, two access control schemes are proposed (MDM–S and MDM–R) and they both achieve higher data average throughput for a given system capacity and outage probability when compared to the persistence state protocol analyzed in [10]. In addition, MDM–R is more efficient in terms of downlink channel utilization. For the case of perfect power control, MDM–S guarantees zero outage probability. Both schemes are easy to implement and do not require exact knowledge of the voice activity parameter. The only restriction is a small slot duration, such that at most ± 1 variation in the number of active voice users from one slot to another can occur. MDM–R allows the fine tuning of system performance to correct the imprecise prior knowledge of some key parameters of the system (e.g. voice activity, standard deviation of SIR, etc.). The protocols have flexible design which permits a natural extension towards more complex data models.

References

[1] A.M. Viterbi, A.J. Viterbi. *Erlang Capacity of a Power Controlled Cellular CDMA System,* IEEE Journal on Selected Areas in Communications, vol 11, No 6, August, 1993, pp. 892-900.

[2] A.J. Viterbi. *Capacity of a Simple Stable Protocol for Short Message Service over a CDMA Network.* Blahut, editor, Communications & Criptography, KAP, 1994, pp. 423-429.

[3] M. Soroushnejad, E. Geraniotis. *Multiple-Access Strategies for an Integrated voice/data CDMA packet radio network,* IEEE Trans. Commun. vol. 43, Feb/Mar/Apr. 1995, pp. 934-945.

[4] A. Sampath, P.S. Kumar, J.M. Holtzman. *Power Control and Resource Management for a Multimedia CDMA Wireless System*, PIMRC'95, September 1995, vol 1., pp 21-25.

[5] N.B. Mandayam, J.M. Holtzman and S. Barberis. *Erlang Capacity for an Integrated Voice/Data CDMA System with Variable Bit-Rate Sources,* PIMRC'95, September 1995, vol. 3, pp. 1078–1083.

[6] N.B. Mandayam, J.M. Holtzman. *Analysis of a Simple Protocol for Short Message Service Data Service in an Integrated Voice/Data CDMA System,* MILCOM'95, San Diego, CA, November 1995, pp. 1160–1164.

[7] M. L. Honig, Joon Bae Kim. *Allocation of DS-CDMA Parameters to Achieve Multiple Rates and Qualities of Service,* Proceeding Globecom 1996, vol 3, pp. 1974-1978.

[8] A. Sampath, N.B. Mandayam, J.M. Holtzman. *Analysis of an Access Control Mechanism for Data Traffic in an Integrated Voice/Data Wireless CDMA System*, VTC'96, vol. 3, pp. 1448-1452.

[9] A. Sampath, N.B. Mandayam, J.M. Holtzman. *Erlang Capacity of a Power Controlled Integrated Voice and Data CDMA System,* VTC'97, vol 3, pp 1557-1561.

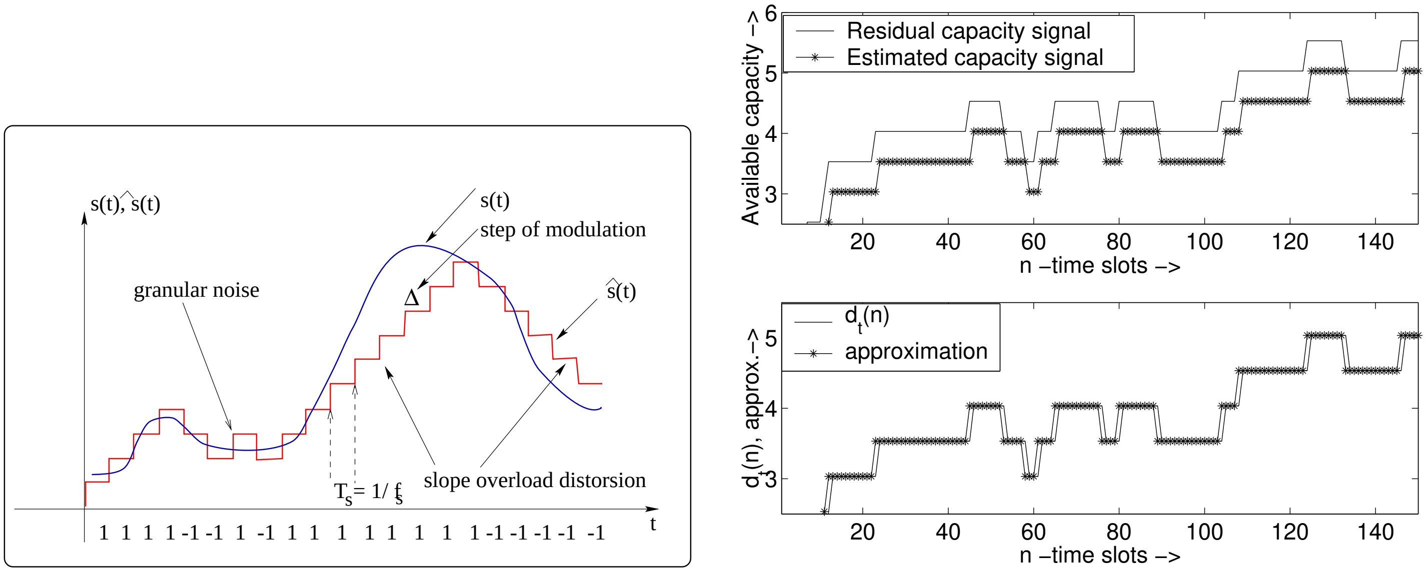
[10] A. Sampath, J.M. Holtzman. *Access Control of Data in Integrated Voice/Data CDMA Systems: Benefits and Tradeoffs,* IEEE Journal on Selected Areas in Communications, vol 15, No 8, October 1997, pp. 1511-1526.

[11] S. Ramakrishna, J. Holtzman *A Scheme for Throughput Maximization in a Dual-Class CDMA System*, IEEE Journal on Selected Areas in Communications, vol 16, No 6, August 1998, pp. 830-844.

[12] T. Liu, J. A. Silvester *Joint Admission/Congestion Control for Wireless CDMA Systems Supporting Integrated Services*, IEEE Journal on Selected Areas in Communications, vol 16, No 6, August 1998, pp. 845-857.

[13] B.P. Lathi. *Modern Digital and Analog Communication systems,* Philadelphia: Holt, Rinehart, and Winston, 1989.

[14] A.J. Viterbi *CDMA: Principles of Spread Spectrum Communication.* Addison-Wesley, 1995



(a) (b)

Figure 1: Delta Modulation: (a) Staircase approximation of an analog signal (b) Proposed Algorithm for Imperfect Power Control (Simulations)

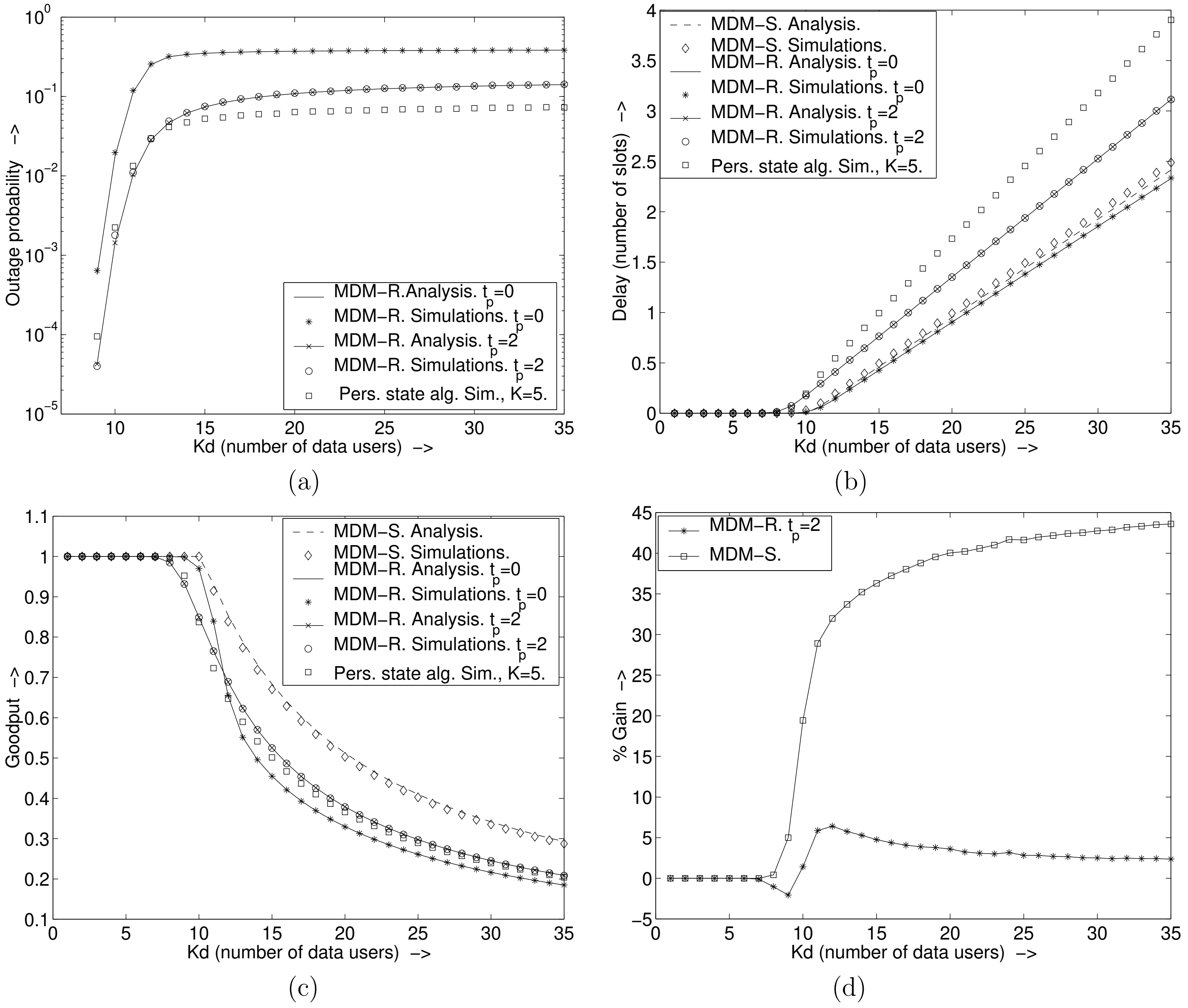


Figure 2: Performance Evaluation for Perfect Power Control: (a) Outage Probability (b) Mean Access Delay. (c) Goodput (d) Percentage gain in average goodput rate for data users

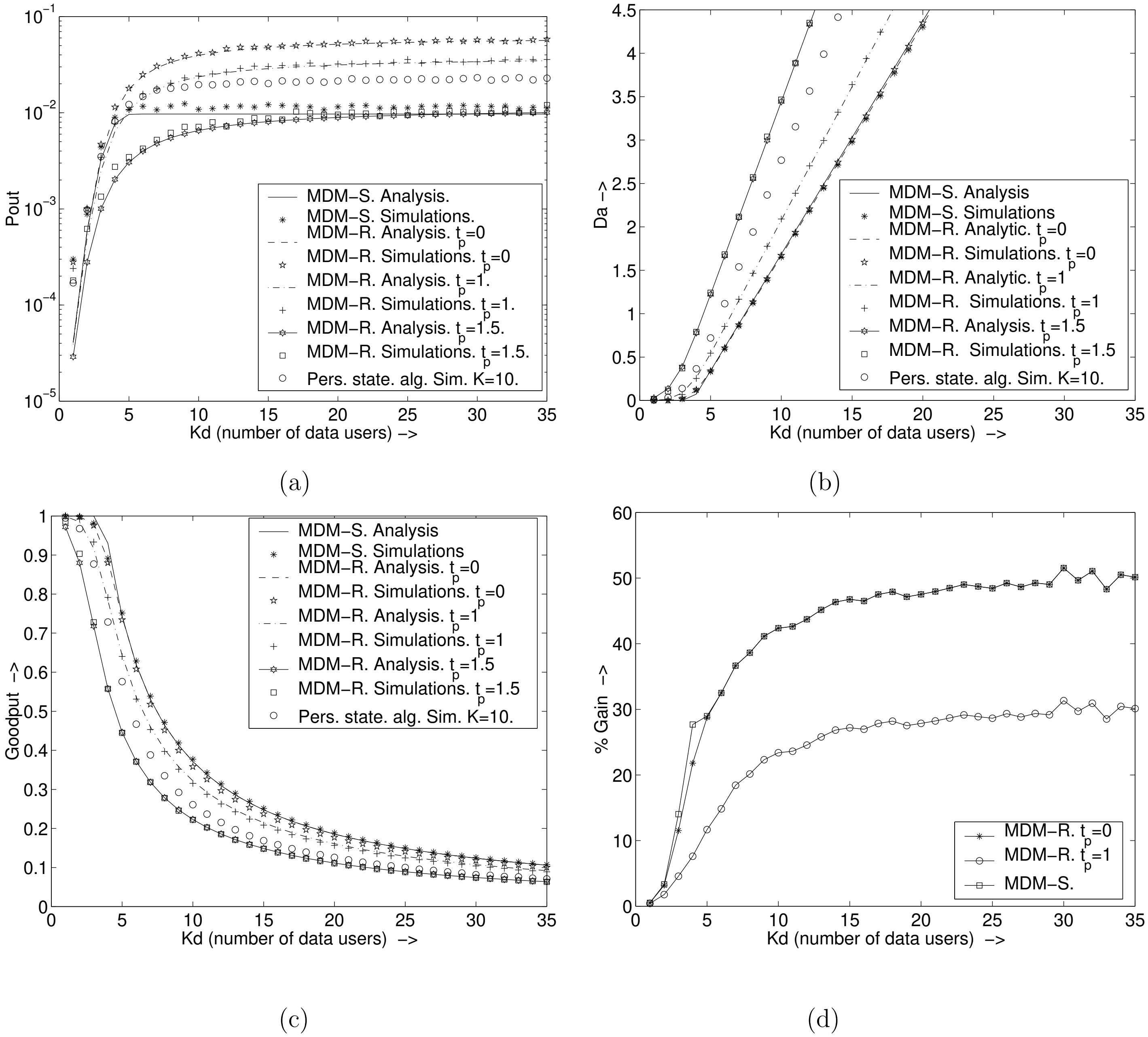
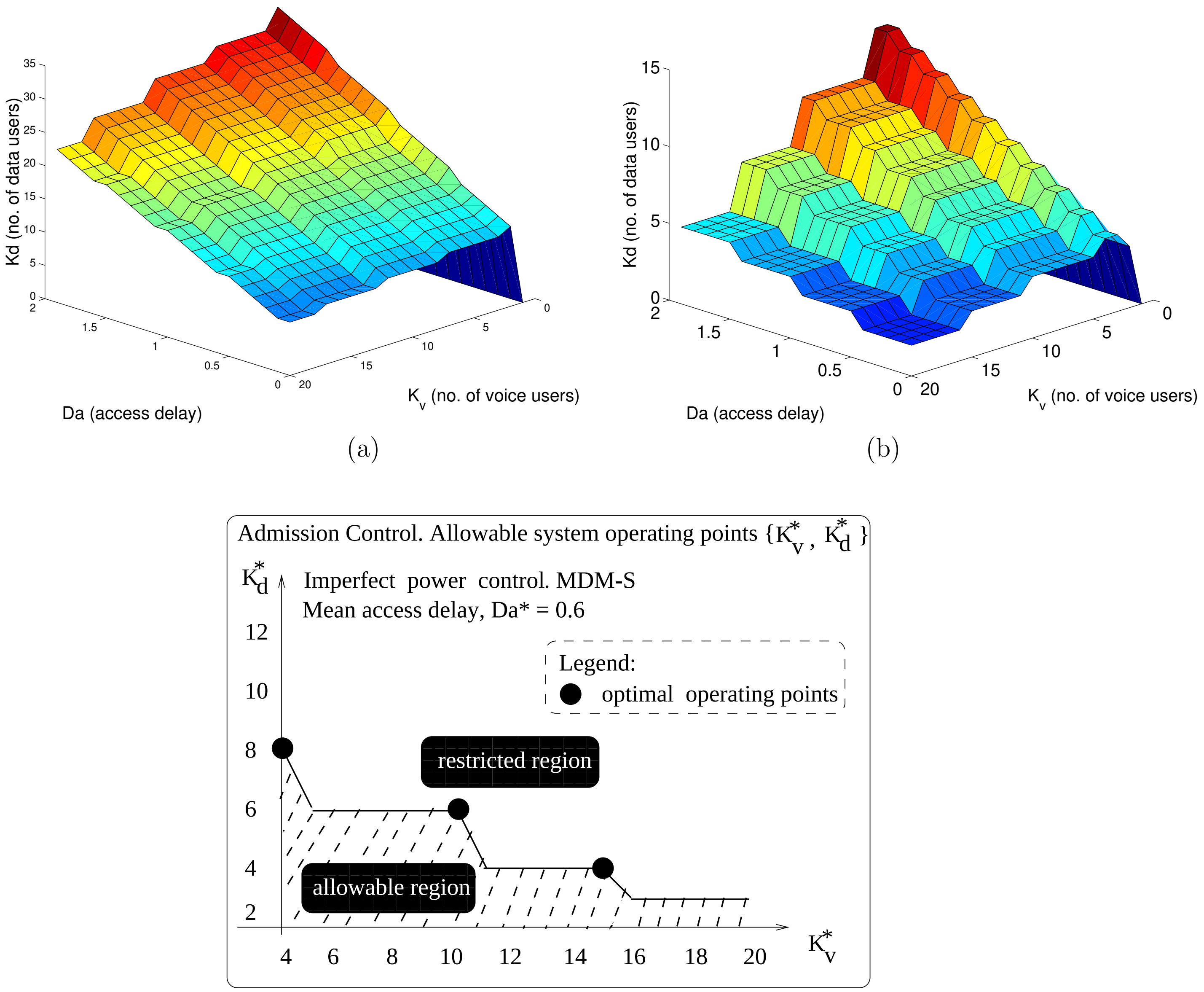


Figure 3: Performance Evaluation for Imperfect Power Control: (a) Outage Probability (b) Mean Access Delay. (c) Goodput (d) Percentage gain in average goodput rate for data users



(c)

Figure 4: Bidimensional system capacity. MDM-S (a) 3D capacity representation for Perfect Power Control (b) 3D capacity representation for Imperfect Power Control (c) Bidimensional capacity – cutting plane through $D_a^* = 0.6$ for the 3D representation. Imperfect Power Control