Optimal Access Control for an Integrated Voice/Data CDMA System

Shruti Mahajan, Manish Singh and Abhay Karandikar

Information Networks Laboratory, Department of Electrical Engineering,
Indian Institute of Technology Bombay, Powai, Mumbai-76, India
karandi@ee.iitb.ac.in

Abstract. Access control (call admission and data transmission policies) becomes essential for optimum resource utilization with traffic having varying Quality of Service (QoS) requirements. In this paper, we determine optimal algorithms for joint admission control of voice calls and transmission of data packets in an integrated voice and data dual-class CDMA system. The objective is to minimize the cost due to blocking of voice calls and the average delay suffered by the data packets. The problem is formulated as an optimization problem over an infinite horizon, and the solution turns out to be a stationary policy for a given set of bit rate, Signal to Interference Ratio (SIR) and outage probability requirements.

1 Introduction

Next generation wireless networks need to be suitably engineered to support a heterogeneous mix of traffic with varying Quality of Service (QoS) requirements. CDMA has been acclaimed as the predominant medium access control (MAC) protocol for next generation systems for its effectiveness in high capacity multi-class traffic systems. The capacity of a CDMA air interface is determined by the capacity of the reverse link, which is typically interference-limited. In this work, we focus on access control of traffic on such a reverse link.

Quite a few attempts have been made to characterize a dual-class CDMA system, with a delay-intolerant voice class and a delay-tolerant data class with higher SIR and bit-rate requirements. Probability of blocking the arriving calls is a direct measure of customer satisfaction for voice subscribers, while the data traffic can be characterized by the queuing delay seen by the packets.

Since the reverse link of a CDMA system is interference limited, the power transmitted by a Mobile Station (MS) is controlled by the Base Station (BS) using sophisticated power control algorithms [1]. In [2], a necessary condition for the feasibility of transmit power assignment for given bit-rate and SIR requirements has been derived. In a perfectly power controlled system, if this condition is not satisfied, SIR demands of all users will not be met, an event termed as *outage*. This condition has been employed by the authors of [3] to schedule data packets in periods of low voice activity. Voice Activity Detection (VAD) capability is assumed to differentiate between voice activity intervals. Though we model our traffic types differently and consider an open system with arrivals and departures, we largely retain their notations for the similarity with their system model and parameters.

The system model in [4] considers two classes (similar to voice and data) of user traffic, with complementary delay tolerance and bit-rate requirements. Their objective is to maximize the throughput of delay tolerant users while minimizing the sum of transmit powers of all mobiles. This work has been extended to include minimization of inter-cell interference in [5]. Both these problems are formulated from a "power allocation" point of view, with a given number of users and do not consider call admission control. For other related work, the readers are referred to [6], [7], [8], [9], [10], [11].

Despite several interesting proposals described above, none of them attempt 'optimal' management of the intrinsic trade-off between admitting more voice calls and delayed scheduling of data packets that are queued up. The primary contribution of our work is to determine a joint admission control and scheduling algorithm which is optimal with respect to the long term average cost due to blocking of voice arrivals and queuing delay of data packets.

2 System Model

We consider a single cell slotted CDMA system. The slot duration is t_0, and is equal to the transmission time of a packet. All arrivals into or departures from the system take place at slot boundaries, and so do any changes in voice activity state of the admitted users.

Voice Traffic There are N voice subscribers, out of which a maximum of M can be supported by the system at any time so as to satisfy the outage probability constraint, which is the probability that target SIR values of all users are not achieved. A voice subscriber who initiates a call and is subsequently accepted into the system is termed as "active". Subscribers not making calls are called "inactive" or "idle". Users which are present in the system just before the start of slot n (where n is the slot index) are denoted by $K_V(n)$. Assuming Voice Activity Detection (VAD), only $V(n)$ of these (active) users might be ON. ON users transmit fixed length packets with a rate of R_V bits/s, OFF users don't transmit (in a practical system, they might transmit at a minimal rate though). We employ geometric probability distributions for the period for which a user remains in the system and for voice activity intervals to model the discrete nature of the system. We define the following probabilities, for events at the start of a slot:

$p =$ probability with which a voice user initiates a call

$q =$ probability with which a voice user departs from the system

$s =$ probability with which a voice user continues to be ON in this slot

$u =$ probability with which a voice user continues to be OFF in this slot

Arrival and departure processes are discrete time batch processes. Thus, at the start of slot n, the number of voice call arrivals, $A_V(n)$, and departures, $G(n)$, are observed. Assume $a_c(n)$ is the number of voice call arrivals accepted, with the decision being instantaneous. These admitted users are considered to be ON users in the current slot n. In addition, users that depart are also considered to be ON in the previous slot. This assumption simplifies our analysis while it introduces a negligible error in the computations.
Users in the system are assumed to change their voice activity state at the beginning of a slot. We denote by $V_{ON}(n)$ the number of voice users that become ON, and by $V_{OFF}(n)$ the number of voice users that become OFF at the start of slot n.

Data Traffic A simple data model with a single queue of maximum length of L packets is considered. Packets of fixed length arrive according to a Poisson process with an average arrival rate of λ_D. A new packet is placed at the end of the queue. It is assumed that the rate is low enough so that the probability of queue length exceeding the maximum length is negligible. Each packet takes exactly one slot to transmit at a rate of R_D bits/s. Multiple packets may be scheduled for transmission on the reverse link in a slot. We assume that r CDMA codes are alloted to the data queue, so that a maximum of r simultaneous transmissions of data packets are possible. This would be possible in a system implementing multi-code CDMA. In case of outage, the transmitted packets are lost. However, the analysis can be easily modified to incorporate the case when the errored packets are retransmitted. Just before the start of slot n, the queue length is denoted by $Q(n)$. At the beginning of the slot, $A_D(n)$ packets arrive, out of which $D(n)$ are scheduled to be transmitted in slot n. Again, the feedback time from BS to the mobile is assumed to be negligible.

An Outage Event Consider the n^{th} slot, i.e., the interval $[n, n+1)$. Just before start of this slot, we have $V(n)$ ON voice users in the system. At the beginning of this slot, $a_c(n)$ new voice users are admitted and $D(n)$ data packets are scheduled for transmission. Considering the departures and the users that change their ON/OFF state (as given in the evolution equation below), we'll now have $V(n+1)$ users that'll be ON during the interval $[n, n+1)$. Suppose the target SIR values for voice and data packets are γ_V and γ_D respectively. With $V(n+1)$ ON voice users and $D(n)$ data packets in transmission in slot n, a feasible power assignment will exist [2] [1] if

$$P\left\{\frac{V(n+1)}{a_V} + \frac{D(n)}{a_D} < 1\right\} \text{where } a_V = W/(R_V\gamma_V) + 1$$
$$a_D = W/(R_D\gamma_D) + 1 \text{ and } W = \text{spread-bandwidth}$$

[1] Note that our definition of variables is different from that of [2]

The event when this condition is violated is termed as outage. The probability of outage is constrained by a given number δ, which is typically a specified system parameter. The constraint

$$P_{out} = P\left\{\frac{V(n+1)}{a_V} + \frac{D(n)}{a_D} > 1\right\} \leq \delta \tag{1}$$

has to be satisfied on a per slot basis. We note here the assumption of perfect power control. For obtaining optimal design constraints, we satisy Equation 1 with equality in our analysis.

The maximum number of voice users that can be admitted into the system, M, can be calculated using the outage condition (1) and the stationary probability distributions of different states of a voice user.

3 Formulation as a Markov Decision Process

The evolution equations for the system variables can be written as

$$K_V(n+1) = K_V(n) + a_c(n) - G(n), \quad Q(n+1) = Q(n) + A_D(n) - D(n)$$
$$V(n+1) = V(n) + a_c(n) - G(n) + V_{ON}(n) - V_{OFF}(n)$$

Define the system state by the 4-tuple $\mathbf{x_n} = \{K_V(n), V(n), A_V(n), Q(n)\}$. The number of states is finite, their number denoted by S. At every state, we take an action, denoted by the 2-tuple $\mu(\mathbf{x_n}) = \{a_c(n), D(n)\}$. The space of all possible actions, $U(\mathbf{x})$, is state-dependent. Given the system is in state $\mathbf{x_n}$ at the observation epoch n (the beginning of n^{th} slot), if action $\mu(\mathbf{x_n}) = \mathbf{u_n}$ is taken, some immediate cost $g_n(\mathbf{x_n}, \mathbf{u_n})$ is incurred (which we quantify later), and the system moves to the next state $\mathbf{x_{n+1}}$ with a probability $p_{\mathbf{x_n}\mathbf{x_{n+1}}}(\mathbf{u_n})$. Such a controlled dynamic system is a Markov Decision Process [12]. The sequence of actions $\Gamma = \{\mu_1, \mu_2, ...\}$ is termed as a *policy*.
Our system satisfies the Markovian assumption, that the next state to be visited depends only on the present state of the system. The transition probability is controlled by taking actions. The transition probabilities are stationary in our problem, which also implies the stationarity of the policy so computed [13].

3.1 State Transitions

Since the ordering among the states does not matter, without loss of generality, we index the states for convenience of notation as follows:
Let $i = \{K_V, V, A_V, Q\}$, $j = \{K'_V, V', A'_V, Q'\}$ and $\mathbf{u} = \{a_c, D\}$.
We also drop the time index n. Then, we have,

$$K'_V = K_V + a_c - G, \quad Q' = Q - D + A_D$$
$$V' = V + a_c - G + V_{ON} - V_{OFF}$$

where, as mentioned earlier, we have assumed that arrivals and departures are all from the pool of ON users. The state transition probabilities can be derived as follows.

$$\begin{aligned}
p_{ij}(\mathbf{u}) &= P\{\mathbf{x_{n+1}} = j | \mathbf{x_n} = i, \mathbf{u_n} = \mathbf{u}\} \\
&= P\{K'_V, V', A'_V, Q' | (K_V, V, A_V, Q), (a_c, D)\} \\
&= P\{K'_V | K_V, a_c\} P\{V' | K'_V, K_V, V, a_c\} P\{A'_V | K'_V\} P\{Q' | Q, D\}
\end{aligned} \quad (2)$$

Now

$$\begin{aligned}
P\{K'_V | K_V, a_c\} &= P\{G = K_V - K'_V + a_c\} \\
&= \binom{V}{G} q^G (1-q)^{V-G}
\end{aligned} \quad (3)$$

Letting $z = V' - V - a_c + G = V' - V + K_V - K'_V$ then

$$\begin{aligned}
P\{V' | K'_V, K_V, V, a_c\} = \sum_{x=0}^{V-G} & \binom{V-G}{x} (1-q-s)^x s^{V-G-x} \\
& \binom{K_V - V}{z+x} (1-u)^{z+x} u^{K_V - V - z - x}
\end{aligned} \quad (4)$$

Voice arrivals have a geometric distribution while the data packets arrive according to a Poisson process. Hence we have,

$$P\{A'_V | K'_V\} = \binom{N - K'_V}{A'_V} p^{A'_V} (1-p)^{N - K'_V - A'_V} \quad (5)$$

$$P\{Q' | Q, D\} = P\{A_D = Q' - Q + D\} = \frac{(\lambda t_0)^{Q' - Q + D} \exp(-\lambda t_0)}{(Q' - Q + D)!} \quad (6)$$

Substituting the numerical values from (3), (4), (5) and (6) in (2), we can compute the transition probability $p_{ij}(\mathbf{u})$.

3.2 Cost Formulation

At every decision epoch, the decision that is taken incurs a cost for the system. The immediate cost is defined as

$$g_n(\mathbf{x_n}, \mathbf{u_n}) = \frac{A_V(n) - a_c(n)}{pN} + \beta \frac{Q(n) - D(n)}{\lambda_D} \quad (7)$$

The first term is the cost incurred due to rejection of voice arrivals, while the latter takes into account the cost due to queuing of data packets. β is a scalar and has been employed to assign desired weights to the two cost terms. This enables us to give priority to voice or data according to the system requirements; e.g. the lower the value of β, the lower will be the fraction of voice calls blocked on arrival. This we verify later through numerical results.

For a finite observation interval, of length say K slots, the total cost incurred will be

$$J_K = g_K(\mathbf{x_K}, \mathbf{u_K}) + \sum_{n=0}^{K-1} g_n(\mathbf{x_n}, \mathbf{u_n})$$

However, the system under consideration can be assumed to evolve over a very long period of time, effectively over an infinite horizon. We seek to minimize the average cost per unit time, defined as [14]

$$\begin{aligned}
&\lim_{K\to\infty} \frac{1}{K}\left\{\sum_{n=0}^{K} g(\mathbf{x_n}, \mu_{\mathbf{n}}(\mathbf{x_n}))\right\} \\
&= \lim_{K\to\infty} \frac{\sum_{n=0}^{K}(A_V(n) - a_c(n))}{KpN} + \lim_{K\to\infty} \beta \frac{\sum_{n=0}^{K}(Q(n) - D(n))}{K\lambda_D} \\
&= \text{blocking probability} + \beta(\text{average delay of data packets})
\end{aligned}$$

where the Little's Theorem has been used for computing the average delay. Thus, we attempt to minimize a function of the long-term blocking probability for voice arrivals and the average delay seen by the data packets.

4 Computing the Optimal Policy

The above optimization problem can be solved using Dynamic Programming techniques. Our system has a large state space, and *Value Iteration Algorithm* is well suited for solving such a system over an infinite horizon. The recursion equation for Value Iteration can be written as

$$J_{k+1}(i) = \min_{\mathbf{u}\in U(i)} \left\{ g(i, \mathbf{u}) + \sum_{j=1}^{S} p_{ij}(\mathbf{u}) J_k(j) \right\}$$

Here i is the current state and $\mathbf{u}$ is one of the feasible actions. After convergence, the equation can be interpreted as follows. The cost of $k+1$ stages is broken into immediate cost and the cost incurred over the next k stages. We note that the time indexing has been reversed here, and $J_k(i)$ can be interpreted as the expected "cost-to-go" with k periods left over the time horizon, when the current state is i and a terminal cost of $J_0(j)$ is incurred when the system ends up in state j. This suggests that for large k, the one-step difference $J_k(i) - J_{k-1}(i)$ will come very close to the minimal average cost per unit time [12].

Let the set of admissible (feasible) policies for a state be denoted by $\Pi(\mathbf{x_n})$. Natural constraints that arise in our system are

$$D(n) \le \min\{Q(n), r\} \text{ and } a_c(n) \le A_V(n)$$

The number of ON voice users along with the number of data packets in transmission should satisfy the outage probability constraint in every slot. For slot n, taking the limiting value of outage probabilty

$$P_{out} = P\left\{\frac{V(n+1)}{a_V} + \frac{D(n)}{a_D} > 1\right\} = \delta \;\Rightarrow\; P\{V(n+1) > C_n\} = \delta \tag{8}$$

$$\text{where } C_n = \lfloor a_V\,(1 - D(n)/a_D)\rfloor$$

Table 1. Values of parameters used for computations

t_0	W	γ_V	γ_D	a_V	a_D	p	q	s	u	δ	λ_D	N	L, r	M	β
20ms	1.25MHz	7dB	10dB	10	7	0.04	0.04	0.67	0.5	0.01	20	15	10	10	0.5

At the start of slot n, before deciding on the action, we know $K_V(n)$, the number of active voice users, and $V(n)$, the number of ON voice users, that were there in the previous slot. Now, for a given value of $a_c(n)$, we can determine the maximum number of data packets that can be scheduled so that P_{out} does not exceed δ.

$$\begin{aligned} V(n+1) &= V(n) + a_c(n) - G(n) + V_{ON}(n) - V_{OFF}(n) \\ \Rightarrow P_{out} &= P\{V_{ON}(n) - V_{OFF}(n) - G(n) > C_n - V(n) - a_c(n)\} \end{aligned}$$

Denoting $K_V(n)$ by K_v and $V(n)$ by V,

$$\begin{aligned} P_{out} = & \sum_{j=C_n}^{K_v+a_c} \sum_{z=0}^{V} \sum_{y=0}^{K_v-V} \sum_{x=0}^{V-z} \binom{V}{z} q^z (1-q)^{V-z} \\ & \binom{K_v - V}{y} (1-u)^y u^{K_v-V-y} \binom{V-z}{x} (1-q-s)^x \\ & s^{V-z-x} \delta_f(j - (V + a_c - z + y - x)) \end{aligned}$$

where j, z, y, x count $V(n+1), G(n), V_{ON}(n), V_{OFF}(n)$ respectively and $\delta_f(x)$ denotes the impulse function. We can thus determine C_n and the corresponding $D(n)$. Let us call this number D_{max}. Accordingly, we modify the admissible values of $D(n)$ to

$$D(n) \le \min\{D_{max}, Q(n), r\}$$

Iterating on the cost J_k, the algorithm searches for the minimum cost policy among the admissible set $\Pi(\mathbf{x})$ for every state $\mathbf{x}$. Let $c_k(\mathbf{x}) = J_{k+1}(\mathbf{x}) - J_k(\mathbf{x})$. The average cost, in the long run, is given by

$$\lim_{k\to\infty} c_k(\mathbf{x}) = \lim_{k\to\infty} (J_{k+1}(\mathbf{x}) - J_k(\mathbf{x}))$$

The convergence criteria used, for a desired level of accuracy ϵ is,

$$\max_{\forall \mathbf{x}} (c_{k+1}(\mathbf{x}) - c_k(\mathbf{x})) < \epsilon \tag{9}$$

The policy so computed is a stationary policy- given that the system is in a particular state, the optimal action is fixed in time. This has a lot of practical utility, since the mapping from states to actions can be computed and stored in the case where traffic characteristics vary slowly with time.

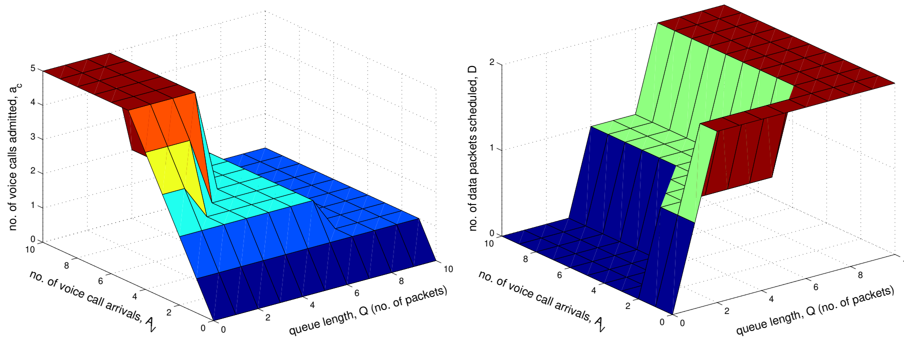

Fig. 1. (a) No. of accepted calls as a function of no. of arrivals and queue length (for $K_V = 5, V = 3$) (b) No. of data packets scheduled as a function of no. of arrivals and queue length (for $K_V = 5, V = 3$)

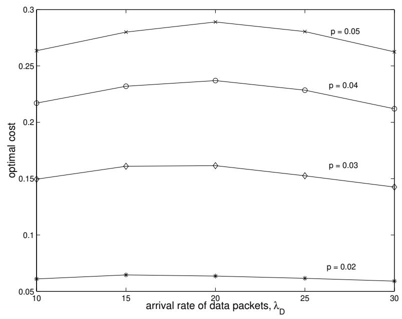

Fig. 2. Optimal cost vs data arrival rate for different values of call initiation probability p

5 Results & Observations

Table 1 gives the values of parameters used in the computations, unless specified otherwise. We consider a representative system with 15 voice subscribers and a data queue with a buffer capacity of 10 packets. The system can support 10 voice users if no data packets were scheduled for a slot. The values of s and u used correspond to mean ON time of 3 slots (60 ms) and mean OFF time of 2 slots (40 ms) respectively. Similarly, $q = 0.04$ corresponds to a mean call holding time of 25 slots, which is 500 ms. The small value of λ_D has been taken so that the average queue length is not more than 5 for a reasonable delay. e.g. for a delay of 5 slots, the average queue length is 2 packets.

In Figure 1(a) and Figure 1(b)we plot the number of accepted calls (a_c) and the number of data packets scheduled (D), respectively, as we vary the number of voice call arrivals (A_V) and the queue length (Q), while the number of users in the system (K_V) and the number of ON voice users (V) are kept constant. The fairness of the optimal policy is evident when, for the same number of

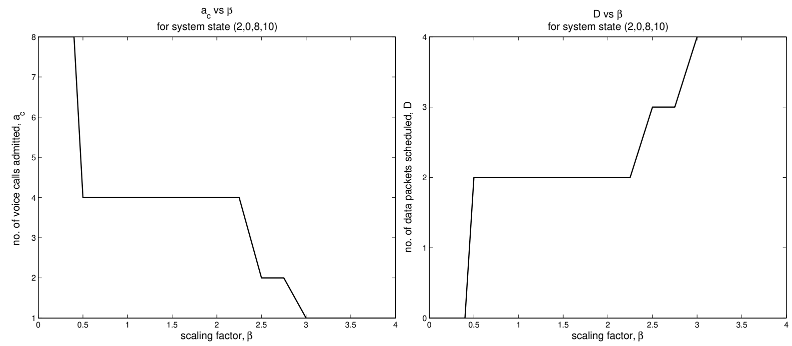

Fig. 3. (a) No. of voice calls accepted vs scaling factor (b) No. of data packets scheduled vs scaling factor

active users, ON users and arrivals in the system, lesser number of voice calls are accepted and more data packets are scheduled as the queue length increases. To take an example, for 5 users in the system out of which 3 are ON, and 5 arrivals, $a_c = 5, D = 0$ when $Q = 3$, while we get $a_c = 2, D = 1$ when $Q = 4$.

On convergence of the algorithm, the differential cost so obtained $(J_{k+1} - J_k)$ is the minimum cost per stage of running the system. We plot the optimal cost per stage versus the parameter λ_D of the Poisson arrival process of data packets. As shown in Figure 2, increasing the probability of voice call initiation (p) results in higher operating costs. This is expected because with increased load on the system, the average queue length and the fraction of rejected voice calls increases and hence the average cost will increase.

The value of β affects the relative priority of voice arrivals or queued data packets. Figure 3 illustrates how the number of accepted voice calls decrease while the number of scheduled data packets increase, respectively, as we increase β, the scaling factor of the data cost term. Values of other parameters are same as given in the table.

6 Conclusions and Future Directions

In this paper, we have considered the problem of admission control of voice users and scheduling (transmission) of data packets on the reverse link of a slotted CDMA system for a single cell scenario. We have computed an optimal policy for access control (voice call admission and data packet transmission) by minimizing a weighted sum of long term blocking probability for voice arrivals and the average delay seen by data packets. The optimization problem has been formulated over an infinite horizon, and solved using Value Iteration. The results illustrate the fairness of such a policy - the data queue is not starved of resources even if the voice load on the system increases. We observe that by increasing the weight of the cost term attributed to the delay of data packets, the number of data packets transmitted (keeping values of other variables same) increases, and vice versa.

In spite of the practical utility of such an algorithm, we note that the solution does not give us actual values of call blocking probability (p_B) and average delay (τ) that are achieved by the system. We can address this shortcoming by solving a constrained optimization problem, that of minimizing p_B while τ is upper bounded by a specified value Υ. The solution thus obtained not only gives us the required policy, but the optimal cost is a lower bound for p_B. Due to lack of space, this formulation has not been discussed here.

References

1. Andrew J. Viterbi *CDMA: Principles of Spread Spectrum Communication* Addison Wesley Longman Inc., 1995
2. Ashwin Sampath, P. Sarath Kumar, Jack M. Holtzman, "Power Control and Resource Management for a Multimedia CDMA Wireless System" *Proceedings of PIMRC'95* Toronto, Canada, September 1995
3. Ashwin Sampath, Jack 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
4. Sudhir Ramakrishna, Jack M. 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
5. Sudhir Ramakrishna, Jack M. Holtzman, "Throughput Maximization in a DS-CDMA System via Time Scheduling with Inter-Cell Interference Constraints" *http://winwww.rutgers.edu/pub/about/people/student/rsudhir/jsac2.ps*
6. Dongxu Shen, Chuanyi Ji, "Admission Control of Multimedia Traffic for Third Generation CDMA Network" *Proceedings of IEEE INFOCOM 2000* Volume: 3, 26-30 Mar 2000
7. C. Comaniciu, N. B. Mandayam, D. Famolari, P. Agrawal, "QoS Guarantees for Third Generation (3G) CDMA Systems via Admission and Flow Control" *Vehicular Technology Conference (VTC), 2000* Volume: 1 , 2000
8. Yue Ma, James J. Han, Kishor S. Trivedi, "Call Admission Control for Reducing Dropped Calls in CDMA Cellular Systems" *IEEE Infocom*, pp.1481-1490, 2000.
9. Mohsen Soroushnejad, Evaggelos Geraniotis, "Multi-Access Strategies for an Integrated Voice/Data CDMA Packet Radio Networks" *IEEE Transactions on Communications* Vol.43, No.2/3/4, February 1995
10. Taekyoung Kwon, Sooyeon Kim, Yanghee Choi, M. Naghshineh "Threshold-type Call Admission Control in Wireless/Mobile Multimedia Networks using Prioritised Adaptive Framework" *Electronics Letters* Volume: 36 Issue: 9 , 27 Apr 2000
11. Jihyuk Choi, Taekyoung Kwon, Yanghee Choi, Mahmoud Naghshineh "Call Admission Control for Multimedia Services in Mobile Cellular Networks: a Markovian Decision Approach" *Proceedings of Fifth IEEE Symposium on Computers and Communications (ISCC), 2000.*
12. Henk C. Tijms *Stochastic Modelling and Analysis- A Computational Approach* John Wiley & Sons
13. Cyrus Derman *Finite State Markovian Decision Processes* Academic Press 1970
14. Dmitri P. Bertsekas *Dynamic Programming and Optimal Control* Vol.I, Athena Scientific, Massachusetts, 1995