Call Admission Control for Multimedia Services in Mobile Cellular Networks : A Markov Decision Approach

Jihyuk Choi
ETRI
Taejon, Korea
choijh@etri.re.kr

Taekyoung Kwon, Yanghee Choi
Seoul National Univ.
Seoul, Korea
{tkkwon, yhchoi}@mmlab.snu.ac.kr

Mahmoud Naghshineh
IBM Watson Research Center
Yorktown Heights, NY 10598
mahmoud@watson.ibm.com

Abstract

There is a growing interest in deploying multimedia services in mobile cellular networks. Call admission control (CAC) is a key factor in quality of service (QoS) provisioning for these services. We propose a CAC which maximizes the revenue by finding out call admission decisions for each state. A semi-Markov decision process is employed to model the cellular system. Also, the QoS requirement such as the handoff dropping probability is taken into consideration. Simulations reveal that our CAC outperforms the upper limit CAC while satisfying the QoS requirements.

1 Introduction

With the recent advances in mobile/wireless communication technologies, there is a growing interest in deploying multimedia services in mobile cellular networks (MCNs). For mobile multimedia services, the existing MCN for voice-oriented services, needs to be adapted in numerous aspects. Call admission control (CAC) is one of such areas in need of adaptation to accommodate multimedia traffic.

The connection-level quality of service (QoS) in MCNs is usually expressed in terms of *call blocking probability* and *call dropping probability* [1, 2]. The call dropping probability is the probability that an accepted call will be forced to terminate before the completion of its service. According to [1, 3], the call dropping probability is directly proportional to the *handoff dropping probability* which is the probability that a handoff attempt fails. Therefore, we focus on the handoff dropping probability as the main QoS requirement in this paper. In addition, an attractive CAC for service providers should maximize the revenue.

For the sake of clarity, multimedia calls belong to multiple and different types of classes and will be refered to as *multiclass* calls henceforth. Typical CAC policies in wireline multiclass networks are *complete sharing* (CS), *complete partitioning* (CP) and *threshold*. In the CS policy, calls of every class share the bandwidth pool. Whereas, in the CP policy, bandwidth for each class is exclusively reserved. In the threshold policy, a newly arriving call is blocked if the number of calls of each class is greater than or equal to a predefined threshold. It is proven that the threshold policy is optimal in two-class networks [4]. However, the threshold policy is not optimal in general multiclass networks.

The above-mentioned three policies are classified as the *coordinate-convex policy*. The coordinate-convex policy boasts of easy tractability; thus, several CAC schemes in this policy have been proposed [4, 7, 8]. However, it turns out that in certain cases coordinate-convex policies are strictly suboptimal. Instead, it is indicated that a CAC using *semi-Markov decision process* (SMDP) can maximize the revenue for multiclass networks [5].

A few proposed CAC policies are based on SMDP. Kwon et al. [9] took into account only one isolated cell and resorted to linear programming (LP) formulation to find out optimal decisions. Yoon et al. [10] also proposed the CAC policy considering a one-cell network; they used an approximation technique to solve an SMDP-formulated problem. However, both approaches take into account only one cell. In this paper, we model a one-dimensional cellular network and describe how to find out optimal admission decisions, which would be useful in more general case.

The rest of this paper is organized as follows. A model of a multiclass MCN is illustrated in Section 2. Section 3 describes how to find out an optimal CAC policy. Numerical results are shown in Section 4. Finally, we conclude this paper in Section 5.

2 Model Description

The cellular system under consideration is one-dimensional (1-D), which is deployed in streets and highways [11]. Our system consists of N cells as shown in Figure 1. Also, we consider a general model of multiclass calls with mobility characteristics.

Suppose that there are K classes of calls in an MCN. Call requests of class-i ($i = 1, 2, ..., K$) in cell-n ($n = 1, 2, ..., N$) are assumed to form a Poisson process with mean arrival rate $\lambda_{n,i}$. The call holding time (CHT) of a class-i call is assumed to follow an exponential distribution with mean $1/\mu_i$. The bandwidth of a class-i call, i.e., the number of channels required to accommodate the call, is denoted by b_i. Finally, for each on-going class-i call, revenue is accrued at rate r_i.

The following simple model is assumed for mobility characterization. A mobile terminal (MT) moves through the whole cellular system. The *cell residence time* (CRT), i.e., the amount of time that an MT stays in a cell before handoff, is assumed to follow an exponential distribution with mean $1/\eta$. Here, we assume that the CRT is independent of class. Hence, any class call in each cell follows the same CRT distribution. Note that the parameter η represents the handoff rate. Moreover, an MT is assumed to handoff to its adjacent cells with equal probability. In our 1-D cellular network, the probability that an MT will handoff to one of its adjacent cells is 0.5. Consequently, the rate that a call in a given cell will handoff to one of its adjacent cells is $\eta/2$.

The total bandwidth (in number of channels) in each cell is the same and denoted by C, assuming a fixed channel allocation (FCA) scheme. Furthermore, it is assumed that the bandwidth of wideband call can be scattered across the bandwidth pool [3, 5, 8].

Our traffic model is summarized in Figure 1. The arrival rate of class-i calls that originate in cell-n is denoted by $\lambda_{n,i}$. The rate of class-i calls that depart from a cell due to service completion is denoted by μ_i. The rate of class-i calls that handoff to our system from outside is denoted by $h_{n,i}$ ($n = 1$ or N). Also, the rate of class-i calls that handoff from a given cell to one of adjacent cells is expressed by $\eta/2$.

The current state of our cellular system is then represented by the vector

$$x = (x_{1,1}, ..., x_{1,K}, x_{2,1}, ..., x_{2,K}, ..., x_{N,1}, ..., x_{N,K}) \quad (1)$$

where $x_{n,i}$ denotes the number of class-i calls in cell-n. In the absence of control, the set Λ of all possible states is given by

$$\Lambda = \{x : \sum_{i=1}^{K} b_i x_{n,i} \le C; x_{n,i} \ge 0; \text{for } n = 1, 2, ..., N\}. \quad (2)$$

Ultimately, for each state x, a CAC policy should find out an "accept/reject" decision for all kinds of traffic (new call and handoff call) in Figure 1.

3 SMDP Approach in Our CAC

3.1 SMDP Overview

The original semi-Markov decision process (SMDP) model [12] considers a dynamic system which, at random points in time, is observed and classified into one of several possible states. After observing the state, a decision has to be made and the corresponding revenue for each state is gained. For each state x, a set of actions is available. This controlled dynamic system is called an SMDP when the following Markovian properties are satisfied: if at a decision epoch the action a is chosen in state x, then the time until, and the state at, the next decision epoch depends only on the present state x.

Consequently, we focus on finding out an optimal decision for each state. An SMDP-formulated problem can be solved by a few methods [12] among which linear programming (LP) has an advantage that additional constraints can be easily incorporated. Thus, using the LP approach, we can guarantee the upper bound of the handoff dropping probability. Finally, we apply the LP approach to solve the SMDP-formulated CAC problem in our cellular system, which aims at both maximum revenue and QoS guarantee.

3.2 Our SMDP Approach

We basically adopt the SMDP approach introduced in [5]. Decisions are made before, rather than after, the occurrence of an event. The SMDP state of the system at a decision epoch is given by the vector $s = (x, e)$, where x is the vector of class-i calls from Section 2. The variable e represents the event type of an arrival and is given by

$$e \in \{a_{1,1}, ..., a_{1,K}, a_{1,K+1}, ..., a_{1,2K}, a_{2,1}, ..., a_{N,2K}\} \quad (3)$$

That is, the indicator $a_{n,i}$ ($a_{n,i} \in \{0, 1\}$,) denotes the origination (event) of a class-i call within the cell-n when $i \le K$, while it denotes the arrival (event) of a class-i call due to handoff from adjacent cells when $i \ge K + 1$.

In this case, when the system is in state x, an accept/reject decision must be made for each type of possible arrival. Thus, the action space B can be expressed by

$$\begin{aligned} B \;=\; & \{(a_{1,1}, a_{1,2}, ..., a_{N,2K}) : a_{n,i} \in \{0,1\} \\ & ; n = 1, 2, ..., N; i = 1, 2, ..., 2K\}. \end{aligned} \quad (4)$$

For example, when $N = 2, K = 2$ and

$$\begin{aligned} & (a_{1,1}, a_{1,2}, a_{1,3}, a_{1,4}, a_{2,1}, a_{2,2}, a_{2,3}, a_{2,4}) \\ & \quad = (1, 0, 1, 1, 1, 1, 1, 1) \end{aligned} \quad (5)$$

then only an origination of a new class-2 call within the cell-1 would be rejected. The action space is actually a state-

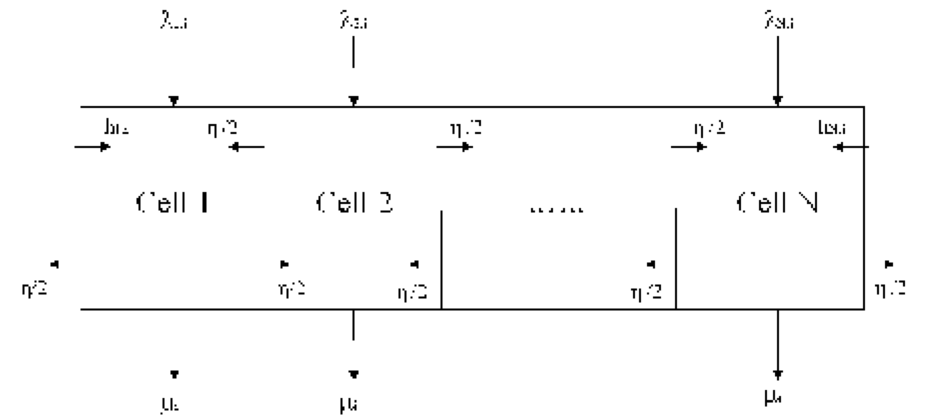

Figure 1. Traffic Model in Our Cellular System.

dependent subset of B denoted by

$$B_x = \{a \in B : a_{n,i} = 0 \; if \; x + e_{n,i} \notin \Lambda\}, \qquad (6)$$

where $e_{n,i}$ is a vector of zeros, except for an one in the $(n*(K-1)+i)$-th position.

If the system is in state $x \in \Lambda$ and the action $a \in B_x$ is chosen, then the expected time, $\tau(x,a)$, until a new state is entered is given by

$$\tau(x,a) = [\sum_{n=1}^{N}\sum_{i=1}^{K}(\lambda_{n,i}a_{n,i} + x_{n,i}\mu_i + x_{n,i}\eta) + \sum_{i=1}^{K}(h_{1,i}a_{1,i} + h_{N,i}a_{N,i})]^{-1}. \qquad (7)$$

On the right hand side of (7), the first two terms, $\lambda_{n,i}a_{n,i}$ and $x_{n,i}\mu_i$, represent the rate at which calls originate from and terminate at cell-n, respectively. The third term, $x_{n,i}\eta$, means the rate of outgoing handoffs from the cell-n, while the last two terms, $h_{1,i}a_{1,i}$ and $h_{N,i}a_{N,i}$, represent the rate of incoming handoffs to the our system. Note that the incoming handoff calls from outside arrive only in the boundary cells (cell-1 and cell-N). To summarize, the transition probability P_{xay} from the state x to any next state $y \in \Lambda$ takes one of the expressions in Table 1.

In the row 1 and 2 of Table 1, $\lambda_{n,i}a_{n,i}$ and $h_{n,i}a_{n,K+i}$ mean the rate of originating of a class-i call and incoming handoffs from the outside of our system, respectively. In the row 3, 4 and 5 of Table 1, $x_{n,i}\mu_i$ and $x_{n,i}\eta/2$ represent the rate of terminating of class-i calls and outgoing handoffs to the outside of our system respectively. In the row 3, 4 and 5 of Table 1, $x_{n,i}(1-a_{n,K+i})\eta/2$ means the rate of dropping incoming handoff class-i calls to the cell-n. Also, In the row 6 and 7 of Table 1 represent the rate of handoffs between cells within our system.

Let $r(x,a)$ be the revenue rate when the cell is in state x and action a has been chosen. If r_i is the revenue rate of class-i call, then the total revenue rate for the cell is calculated by

$$r(x,a) = \sum_{i=1}^{K} r_i x_i. \qquad (8)$$

The utilization of the system can be obtained by replacing r_i with b_i in (8). Precisely, we should also consider the effect of handoff call dropping into the definition of revenue rate. However, we assume that the effect of bandwidth used by forced-terminated calls is negligible in our LP formulation.

3.3 LP Formulation

The LP associated with our SMDP for the maximum revenue is given by (9), where the decision variable z_{xa} represents "the long-run fraction of decision epochs at which the system is in state x and action a is taken."

$$\begin{aligned} \text{maximize} \; & \sum_{x\in\Lambda}\sum_{a\in B_x} r(x,a)\tau(x,a)z_{xa} \\ \text{s.t.} \; & \sum_{x\in\Lambda}\sum_{a\in B_x} \tau(x,a)z_{xa} = 1, \\ \sum_{a\in B_y} z_{ya} = & \sum_{x\in\Lambda}\sum_{a\in B_x} P_{xay}z_{xa}, \; y \in \Lambda, \\ & z_{xa} \geq 0, \; x \in \Lambda, \; a \in B_x. \end{aligned} \qquad (9)$$

By solving the LP formulation in (9), we can obtain an optimal CAC policy. However, we also need to consider the QoS requirements: the upper bound of the handoff dropping probability. Let D_i denote the maximum tolerable handoff dropping probability of a class-i call. Here, we should consider this probability from two aspects: external handoff

Table 1. Transition Probability P_{xay}

1	$(\lambda_{n,i}a_{n,i} + h_{n,i}a_{n,K+i})\tau(x,a)$	$y = x + e_{n,i}$,	$n = 1$ or N,
2	$\lambda_{n,i}a_{n,i}\tau(x,a)$	$y = x + e_{n,i}$,	$n = 2, ..., N-1$,
3	$x_{n,i}(\mu_i + \eta/2 + (1 - a_{2,K+i})\eta/2)\tau(x,a)$	$y = x - e_{n,i}$,	$n = 1$,
4	$x_{n,i}(\mu_i + (1 - a_{n-1,K+i})\eta/2 + (1 - a_{n+1,K+i})\eta/2)\tau(x,a)$	$y = x - e_{n,i}$,	$n = 2, .., N-1$,
5	$x_{n,i}(\mu_i + (1 - a_{N-1,K+i})\eta/2 + \eta/2)\tau(x,a)$	$y = x - e_{n,i}$,	$n = N$,
6	$x_{n,i}\ a_{n+1,K+i}\ \eta/2\ \tau(x,a)$	$y = x - e_{n,i} + e_{n+1,i}$,	$n = 1, ..., N-1$,
7	$x_{n,i}\ a_{n-1,K+i}\ \eta/2\ \tau(x,a)$	$y = x - e_{n,i} + e_{n-1,i}$,	$n = 2, ..., N$,
8	0	otherwise	

from outside and internal handoff between cells in our system. Then the following constraint needs to be added to the LP formulation in (9) to satisfy the QoS requirements:

$$\sum_{x\in\Lambda}\sum_{a\in B_x}(1 - a_{n,K+i})\tau(x,a)z_{xa} \le D_i, \quad \text{for } i = 1,2,...,K \text{ and } n = 1 \text{ or } N \tag{10}$$

$$\sum_{x\in\Lambda}\sum_{a\in B_x}[D_i(\sum_{n=1}^{N-1}x_{n,i} + \sum_{n=2}^{N}x_{n,i}) -\{\sum_{n=1}^{N-1}(1 - a_{n+1,K+i})x_{n,i} + \sum_{n=2}^{N}(1 - a_{n-1,K+i})x_{n,i}\}]\tau(x,a)z_{xa} \ge 0, \quad \text{for } i = 1,2,...,K. \tag{11}$$

Equation (10) represents the upper bound of the handoff dropping probability from outside. Recall that the handoff dropping probability is the probability that handoff attempt fails. Then, the handoff dropping probability between cell-$(n-1)$ and cell-n is given by

$$\frac{\tau(x,a)z_{xa}\{(1 - a_{n,K+i})x_{n-1,i}\ \eta/2 + (1 - a_{n-1,K+i})x_{n,i}\ \eta/2)\}}{\tau(x,a)z_{xa}(x_{n-1,i}\ \eta/2 + x_{n,i}\ \eta/2)}. \tag{12}$$

Next, we can extend (12) to the all cells within our system. The upper bound of the handoff dropping probability between cells is represented by

$$(\sum_{x\in\Lambda}\sum_{a\in B_x}\tau(x,a)z_{xa}\{\sum_{n=1}^{N-1}(1 - a_{n+1,K+i})x_{n,i}\ \eta/2 + \sum_{n=2}^{N}(1 - a_{n-1,K+i})x_{n,i}\ \eta/2\}) \le D_i(\sum_{x\in\Lambda}\sum_{a\in B_x}\tau(x,a)z_{xa}\{\sum_{n=1}^{N-1}x_{n,i}\ \eta/2 + \sum_{n=2}^{N}x_{n,i}\ \eta/2\}), \quad \text{for } i = 1,2,...,K. \tag{13}$$

Table 2. Complexity of the Simulation (C=5, K=2, b_1=1, b_2=2).

N	rows	columns	non-zeros
1	16	110	12
2	150	12320	144

¿From (13), we can get the constraint (11).

4 Numerical Results

For numerical results, we simulated one-cell model ($N = 1$) and two-cell model ($N = 2$) because the complexity increases exponentially as N increases. Table 2 illustrates the complexity of two models.

We compare our SMDP CAC with the upper limit (UL) CAC policy [13] that has a threshold t_i for a class-i call originating in a cell. (The UL CAC is equivalent to the threshold CAC.) For example, for a two-class case in one cell, the UL policy with threshold (2,1) blocks a new class-1 call originating in a cell if there are already at least two class-1 calls in the cell. Similarly, it blocks a new class-2 call originating in the cell if there is already at least one class-2 call in the cell. An incoming handoff call of any class is accepted only if there are enough available channels for the call.

As the number of bandwidth units (channels) increases, the complexity of the corresponding LP formulation increases exponentially [5]. Hence, we compare the above CAC policies in the case of small bandwidth capacity. Here, We let $C = 5$, $K = 2$, $b_1 = 1$, $b_2 = 2$, $D_1 = 0.02$ and $D_2 = 0.04$.

Simulations are carried out as the Erlang load ($\lambda_{n,i}/\mu_i$) of every class increases. Moreover, the handoff call arrival rate to our system is assumed to be proportional to the new call arrival rate by $h_{n,i} = \alpha\lambda_{n,i}$ ($n = 1$ or N) for every class. Here α is set to 0.5. Throughout the experiments, the CHT ($1/\mu_i$) and the CRT ($1/\eta$) is assumed to follow expo-

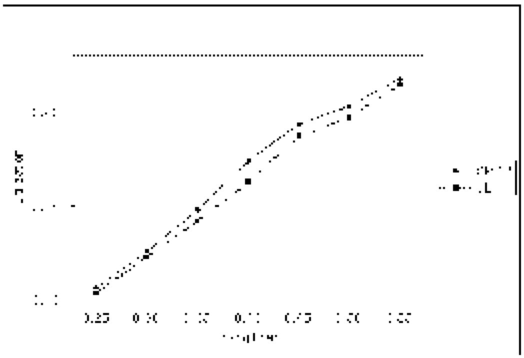

Figure 2. Utilization vs. Erlang Load (N=1).

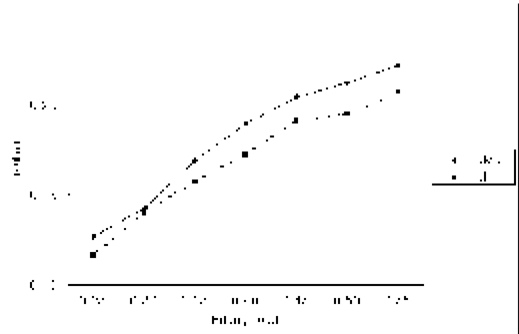

Figure 3. Utilization vs. Erlang Load (N=2).

nential distribution with mean 100 seconds and 60 seconds, respectively [14].

The utilization of a cell is the main performance metric in the above simulation experiments. The utilization of both CAC policies in the case of one-cell model ($N = 1$) and two-cell model ($N = 2$) are shown in Figure 2 and Figure 3. We adopt the highest utilization of UL policy that satisfies the upper bound of handoff dropping probability. Figure 4 , Figure 5 and Figure 6 demonstrate that our policy and UL policy satisfy the upper bound of handoff dropping probability.

Figure 7 shows the *revenue ratio* in the cases of one-cell model and two-cell model. Here, the revenue ratio is defined as the ratio of the revenue of our SMDP policy to that of UL policy. As shown in Figure 7, the revenue ratio of two-cell model is generally greater than that of one-cell model. We believe that the advantage of our policy can increase accordingly as the number of cells in our model increases.

Note that solving LP problem is an off-line procedure. That is, call admission decisions are found out a priori before operating CAC. Also, the techniques for solving large-scale LP problem such as [15] can be used to apply to the

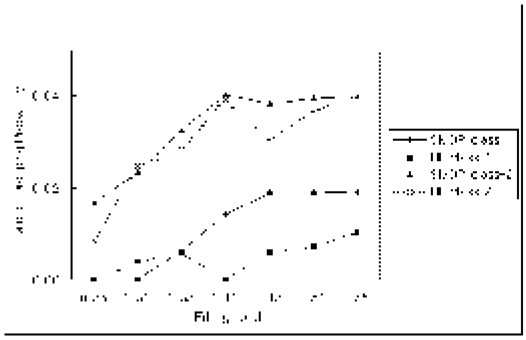

Figure 4. Handoff Dropping Probability from the outside vs. Erlang Load ($N = 1$).

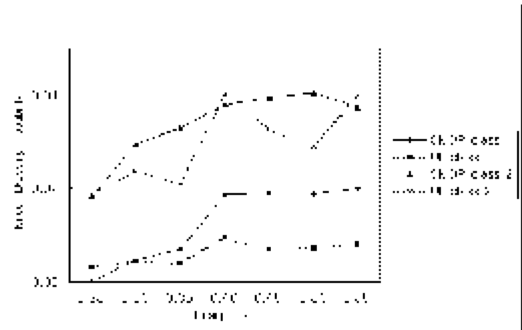

Figure 5. Handoff Dropping Probability from outside vs. Erlang Load ($N = 2$).

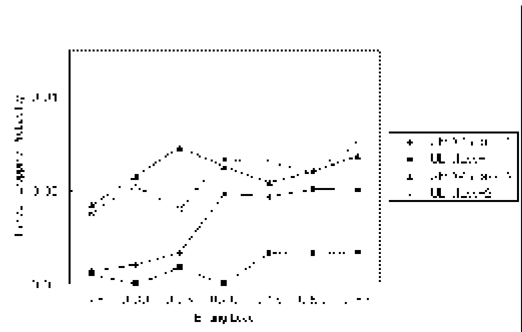

Figure 6. Handoff Dropping Probability between Cells vs. Erlang Load ($N = 2$).

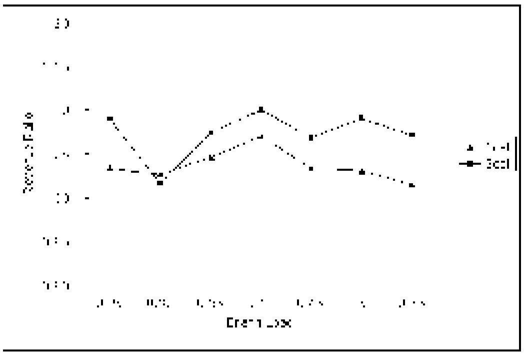

Figure 7. Revenue Ratio vs. Erlang Load.

cases of large cellular systems and/or cellular systems with large capacity.

5 Conclusion

It is anticipated that demands for multimedia services will grow in future mobile cellular networks. Optimal CAC is essential for the efficient utilization of scarce radio bandwidth. We proposed an optimal CAC policy that can maximize the revenue while satisfying the QoS requirements. The proposed CAC policy models our cellular systems using the semi-Markov decision process (SMDP). The linear programming method for solving the SMDP is employed to find out the optimal CAC decisions for each state. Simulation experiments are carried out to show that our CAC outperforms the upper limit CAC while satisfying the upper bound of handoff dropping probability.

References

[1] Stephen S. Rappaport, "The Multiple-Call Hand-off Problem in High-Capacity Cellular Communications Systems," *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 3, pp. 546-557, Aug. 1991.

[2] Mahmoud Naghshineh and Mischa Schwarz, "Distributed Call Admission Control in Mobile/ Wireless Networks," *PIMRC '95*, pp. 289-293, 1995.

[3] Brocha Epstein and Mischa Schwarz, "QoS Based Predictive Admission Control for Multi-media Traffic," *9-th Tyrrhenian Workshop on Digital Communications*, Lerici, Italy, Sep. 1997.

[4] Keith W. Ross and Danny H. K. Tsang, "The Stochastic Knapsack Problem," *IEEE Transactions on Communications*, Vol. 37, No. 7, pp. 740-747, July 1989.

[5] Keith W. Ross and Danny H. K. Tsang, "Optimal Circuit Access Policies in an ISDN Environment: A Markov Decision Approach," *IEEE Transactions on Communications*, Vol. 37, No. 9, pp. 934-939, Sep. 1989.

[6] Keith W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*, Springer, 1995.

[7] Scott Jordan and Pravin P. Varaiya, "Control of Multiple Service, Multiple Resource Communication Networks," *IEEE Transactions on Communications*, Vol. 42, No. 11, pp. 2979-2988, Nov. 1994.

[8] Chi-chao Chao and Wai Chen, "Connection Admission Control for Mobile Multiple-class Personal Communication Networks," *IEEE Journal on Seleted Areas in Communications*, Vol. 15, No. 8, pp. 1618-1626, 1997.

[9] Taekyoung Kwon,Yanghee Choi and Mahmoud Naghshineh, "Optimal Distributed Call Admission Control for Multimedia Services in Mobile Cellular Network," *MoMuc '98*, pp. 263-266, 1998.

[10] In-Soo Yoon and Byeong Gi Lee, "A Distributed Dynamic Call Admission Control that Supports Mobility of Wireless Multimedia Users," *ICC '99*, pp. 1442-1446, 1999.

[11] David A. Levine, Ian F. Akyildiz and Mahmoud Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept," *IEEE/ACM Transactions on Networking*, Vol. 5, No. 1, pp. 1-12, Feb. 1997.

[12] Henk C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*, John Wiley and Sons, 1986.

[13] Subir K. Biswas and Bhaskar Sengupta, "Call Admissibility for Multirate Traffic in Wireless ATM Networks," *Infocom '97*, pp. 650-658, 1997.

[14] Sirin Tekinay and Bijan Jabbari, "A Measurement-Based Prioritization Scheme for Handovers in Mobile Cellular Network," *IEEE Journal on Seleted Areas in Communications*, Vol. 10, No. 8, pp. 1343-1350, 1992.

[15] G. Y. Zhao, "Interior-Point Methods with Decomposition for Solving Large-Scale Linear Programs," *Journal of optimization theory and applications*, Vol. 102, No.1, pp. 169 - 192, 1999