Optimal CAC in MultiService Resource Sharing System with Preemptive Priority

Glaucio H.S. Carvalho, Rodolfo W.L. Coutinho, João C. W. A. Costa.

Resumo— Uma análise de um sistema de compartilhamento de recursos multiserviço é apresentada neste trabalho. A principal contribuição é o projeto de um Controle de Admissão de chamadas (CAC) ótimo que considera o impacto do conjunto de atividades envolvidas em uma operação de preempção em sua decisão ótima, o que até então não foi estudo na literatura. Nós usamos o processo Semi-Markoviano de Decisão para modelar o comportamento do sistema. Nós mostramos por meio de uma análise das medidas de Qualidade de Serviço (QoS) que o desempenho do CAC ótimo é fortemente afetado quando os custos relacionados com a operação de preempção é variado.

Palavras-Chave— Controle de Admissão de Chamadas, Sistema de Compartilhamento de Recursos, Prioridade Preemptiva.

Abstract— An analysis of multiservice resource sharing system with preemptive priority is presented in this paper. The main contribution is the design of an optimal Call Admission Control (CAC), which considers the impact of the set of activities involved in preemption operation on its optimal decision, which has not yet previously studied in literature. We use a Semi Markov Decision Process for modeling the system behavior. We show by means of an analysis of the Quality of Service (QoS)performance measurements that the optimal CAC performance is greatly affected when the costs associated with the preemption operation is varied.

Keywords— Call Admission Control, Resource Sharing System, Communication Network, Preemptive Priority.

I. INTRODUCTION

Resource sharing is one of the main characteristics in communication and networking systems. In this context, Call Admission Control (CAC) plays a crucial role by determining how the system resources should be shared among the different service classes in order to achieve good performance. The design of CAC has been subject of many studies in literature; for instance, for wireless networks (see [1], [2]) and wired network (see [3], [4]). Aiming at optimizing the resource allocation, optimality studies have also been conducted. The bulk of them uses the Semi-Markov Decision Process (SMDP) framework as the mathematical tool for modeling and solving the optimal control problem [5], [6],[7],[8],[9],[10],[11].

In multiservice resource sharing systems like ATM/DSL/WiMax/PLC/Cellular mobile networks, different service classes must be handled accordingly their Quality of Service (QoS) requirements. For instance, real time traffic is delay-sensitive and tolerant to a certain level of loss. On the other hand, non real time traffic is often delay-insensitive

Glaucio Haroldo Silva de Carvalho. Faculdade de Estatística, Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, Brasil. Rodolfo W.L. Coutinho, João Crisóstomo Weyl A. Costa, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, Brasil, E-mails: {ghsc,rwlc,jweyl}@ufpa.br. Este trabalho foi parcialmente financiado pelo CNPq.

and tolerant to transmission rate variations, but requires reliable transmission. A way to keep the high performance of high priority service classes over low priority service classes is by means of preemptive priority. Accordingly [12], the preemptive priority allows a high priority service class being served as soon as it arrives into the system. Hence, since this priority provides resource assurance, it has been studied in different contexts in communication and networking systems. For example, in wireless networks, the preemptive priority is often employed to keep the high performance of voice traffic [14]-[18]. In [19], it is studied the re-allocation of radio channels between voice and data calls upon voice call arrival at the GSM/GPRS networks. The re-allocation happens when there are no free radio channels to accommodate an incoming voice request. In this case, one slot of a multi-slot GPRS call is de-allocated and allocated to the voice call request. In [20], the preemptive priority is used to re-allocate radio channels between slow and fast moving users, upon handoff fast user´s call arrival at the mobility-aware CAC. In Wavelength-Division Multiplexing (WDM) network, the preemptive priority is given for circuits over the bursts in the hybrid optical switches inside the WDM core networks [22][23]. Recently, preemptive priority has been employed to in favor of real time service in the traffic between Internet Service Provider and Home Area Networks (HAN) [21].

During a preemption occurrence three events take place: *i) de-allocation and allocation to an incoming high priority service class an amount of bandwidth used by a low priority service class; ii) low priority service classes Quality of Service (QoS) degradation and/or the low priority service classes dropping; iii) additional signaling overhead is transmitted to run and manager the preemption operation in the link in which it occurs as well as in others where the extra signaling traffic is carried out.* These events are not mutually exclusive; actually, they always occur together.

In the design of CAC with preemptive priority, the previous works consider events *i* and *ii* in their modeling process without any cost associated with them in order to identify their impact of the system performance; in the design of optimal CAC, the previous works frequently seek for an optimal admission policy. In this paper, we propose and optimal CAC, which looks for an optimal admission and preemption policy. To the best of our knowledge, we have not found any work is literature that address the problem of optimal CAC in multiservice resource sharing systems with preemptive priority. Also, we have not found any work that looks for an optimal admission and preemption policy.

The CAC proposed aims at minimizing high priority service classes blocking probabilities, taking low service classes pre-

emption into account. Without losing generality, it is assumed that high service classes are real time calls and low service classes are non real time calls. In order to optimally control the use of the preemptive priority, we include in the cost function two costs: the preemption cost and overhead cost. Theses costs are functions of the number of preempted non real time connections and the amount of bandwidth rearranged or moved by the preemption operation, respectively. They refer to the event *i* and *ii* stated above. The event *iii* is not taking into account in this paper; however, there is not difficulty to include it in this analysis. For example, see paper [6] in which a signaling cost associated with signaling overhead and processing load incurred when vertical handoff execution is performed. Also, in order to mitigate the impact of the preemptive priority on non real time services, we use the degradation and compensation mechanism to capture the elastic characteristic of non real time. The degradation process means that accordingly the network dynamic, the bandwidth allocated to an ongoing call may be gradually reduced, while compensation consists of the reverse process [13], [14]. We solve the optimal control optimization by using the SMPD framework and compute the optimal admission and preemption policy by using the value iteration algorithm.

The rest of the paper is organized as follows: Next Section introduces the network model and describes in detail the SMDP modeling approach as well as the performance measurements. Section III deals with the numerical results. Finally, conclusions are drawn in in Section IV.

II. THE MODEL AND THE OPTIMIZATION PROBLEM

A. Traffic Assumptions

The system under consideration consists of a link with B resources, which is shared by real time connections and non real time connections. As usual, a real time connection is delay-sensitive and requires a constant bandwidth in order to meet its QoS requirements. On the other hand, a non-real time connection has less stringent requirements, *i.e.*, as an elastic traffic, it can tolerate variations in the service rate thanks to the TCP flow control mechanism. In this sense, non-real time connections equally share the resources not used by real time connections, which means that the each non-real time connection service rate can change over time, depending on the number of ongoing real time connections and non real time connections.

For the sake of Markov modeling, a real time service class i (among the N real time service classes) arrives according to a Poisson process with parameter λ_{ri}. Each real time connection request demands B_i resources and requires negative exponential service time with mean rate $1/\mu_{ri}$. In this paper, we consider that all the non real time traffic are aggregated in an unique service class. This is because the degradation and compensation mechanism is capable of covering a lot of different data applications [13][14]. Thus, the non real time service connection arrives according to a Poisson process with parameter λ_{nr}. Each non real time connection request can adjust its bandwidth in the range of values $[B_{min}, B_{max}]$ resources and *ideally* requires negative exponential service

time with mean rate $1/\mu_{nr}$. Finally, we define $\rho_{ri} = \lambda_{ri}/\mu_{ri}$ ($0 \le i \le N$) and $\rho_{nr} = \lambda_{nr}/\mu_{nr}$ as the real time service class i and non real time connections intensities, respectively.

B. Space State

We define in Eq.(1) the set Φ of all feasible states in which m_{ri} and m_{nr} are the number of ongoing real time service class i connections and non-real time connections, respectively. Since a real time service class i demands B_i resources to fulfill its QoS profile, its maximum number of connections in the system is given by $\lfloor \frac{B}{B_i} \rfloor$, where $\lfloor g \rfloor$ is the largest integer not greater than g. Likewise for non real time service, but its maximum number of connection in the system is $\lfloor \frac{B}{B_{min}} \rfloor$.

In our proposal, we use the preemptive priority to give resource assurance for real time services over non real time services. In order to mitigate the preemption impact on the network performance, the elastic characteristic of non real time traffic is taking into account. In doing so, we use the degradation and compensation mechanism. This way, a non real time connection can finish its service quickly by using whenever possible high amounts of bandwidth.

The modus operandi of the degradation and compensation mechanism assume that whenever possible a non real time connection is accepted and served with the maximum bandwidth B_{max}; however, due to the resource dynamic occupancy, it will adjust the actual bandwidth values between the minimal bandwidth B_{min} and the maximum bandwidth B_{max} after any system state changes motivated by call arrivals or departures. To model this traffic elasticity, it is used the concept of ideal departure rate, [13], [14], in which the real instantaneous departure rate of data connections is proportional to the actual bandwidth of each connection. So, with N real time service classes into the system, each non real time connection will receive the bandwidth of

$$b_w(x) = min(B_{max}, max(1, \frac{B - \sum_{i=1}^{N} m_{ri}B_i}{m_{nr}})), \quad (2)$$

whether $0 \le \sum_{i=1}^{N} m_{ri}B_i < B, m_{nr} > 0, x \in \Phi$; and will be served with service rate of

$$\mu_{nrx} = \frac{b_w(x)}{B_{max}}\mu_{nr}, \ x \in \Phi. \quad (3)$$

It is worthy to note that inside the concept of *ideal departure rate* when a non real time connection receives the maximum bandwidth, B_{max}, its mean service rate will also be maximized and equal to $\mu_{nrx} = \mu_{nr}$.

The random variable e, in Eq.(1), is the last event occurred. This information is introduced in the state space in order to define the set of possible actions in each state. Accordingly the system dynamics, the values of e may be either 0 or $1 \le i \le N$; where the former means the arrival (departure) of a non real time connections or departure of a real time service class i connection and latter means an arrival of a real time service class i.

$$\Phi = \{(m_{r1}, m_{r2}, \cdots, m_{nr}, e) / \sum_{i=1}^{N} m_{ri}B_i + m_{nr}b_w(x) \le B, x \in \Phi; e \in [0, 1, 2, \cdots, i, \cdots, N-1, N]\} \qquad (1)$$

C. Decision Epochs and Actions

We assume that each state means the system's configuration just after an event occurrence and just before a decision making. The "real" decision epochs are the arrivals of real time connections, i.e., $e = 1, 2, \cdots, N$; while the service completion epochs and arrival of non real time connection are defined as "fictitious" decision epochs, $e = 0$. In each state $x \in \Phi$, the admission controller can choose one out of the possible actions:

$$A(x) = \begin{cases} a = 0, & 0 \le e \le N; \\ a = 1, & 1 \le e \le N; \sigma \le B; \\ a = 2, & 1 \le e \le N; (\sigma \le B) \text{ or} \\ & (\sigma > B \text{ and } m_{nr}b_w(x) \ge B_j); \end{cases}$$

where $\sigma = B_j + \sum_{i=1}^{N} m_{ri}B_i + m_{nr}b_w(x)$ is the amount of bandwidth used by real time connections and non real time connections plus the bandwidth required by the new real time connection request of type j $(1 \le j \le N)$. In the set of actions $a \in A(x), x \in \Phi$, the action $a = 0$ denotes the rejection, $a = 1$ denotes acceptance and $a = 2$ denotes acceptance with preemption.

Since there is a minimum bandwidth requirement for non real time connections, it is needed to determine if the remainder bandwidth is enough to accommodate all the existing ones into the system. Thus, after the admission with preemption, the remainder bandwidth can support $\theta = \lfloor \frac{B - \sum_{i=1}^{N} m_{ri}B_i}{B_{mim}} \rfloor$ non real time connections with bandwidth B_{mim}. Thus, whether $m_{nr} < \theta$, then the system can support all the existing non real time connections with bandwidth more than B_{mim}; otherwise, $\zeta = m_{nr} - \theta$ non real time connections will be preempted and the system will reduce the bandwidth of the remainder (θ) to B_{mim}. Consequently, the number of non real time connections into the system after the admission will be given by $min(m_{nr}, \theta)$.

D. Expected Time Until the Next Decision Epoch

If the system is in the state $x \in \Phi$ and the action $a \in A(x)$ is chosen, then the expected time until the next decision epoch, $\tau_x(a)$, is given by:

$$\tau_x(a) = \frac{1}{\sum_{i=1}^{N} \lambda_{ri} + \sum_{i=1}^{N} m_{ri}\mu_{ri} + \lambda_{nr} + m_{nr}\mu_{nrx}}. \qquad (4)$$

E. State Dynamics

The state dynamic is completely specified by stating the transition probabilities among the system states. Thus, let $p_{xy}(a)$ be the probability that in the next decision epoch the state will be $y \in \Phi$ if the present state is $x \in \Phi$ and the action $a \in A(x)$ is chosen. For all x and $y \in \Phi$, we have the cases presented in Eq.(5):

F. Cost Function

If the system is in the state $x \in \Phi$ and the action $a \in A(x)$ is chosen, the admission control incurs in the following cost

$$C_x(a) = C_b(x, a) + C_p(x, a) + C_o(x, a), \qquad (6)$$

where $C_b(x, a)$, $C_p(x, a)$, and $C_o(x, a)$ are the real time service class connection blocking cost, the preemption cost, and the overhead cost, respectively. The former is given by

$$C_b(x, a) = c_{bi}, \qquad x \in \Phi, 1 \le e \le N, a = 0 \in A(x), \quad (7)$$

where c_{bi} is the real time service class i connection blocking cost incurred whenever a connection request that belongs to this service class is blocked.

The preemption cost is computed as

$$C_p(x, a) = \zeta c_p, \qquad x \in \Phi, 1 \le e \le N, a = 2 \in A(x), \zeta > 0, (8)$$

where c_p is the immediate cost incurred whenever an incoming real time connection is accepted and one non real time connection is preempted.

Let m'_{nr} and m''_{nr} be the number of ongoing non real time connections just before and after a decision making. Also, let $y \in \Phi$ be the system state just after a decision making. We define the overhead cost as

$$C_o(x, a) = (m'_{nr}b_w(x) - m''_{nr}b_w(y))c_o, \qquad (9)$$

with $x \in \Phi, 1 \le e \le N, a = 2 \in A(x)$; where c_o is the immediate cost incurred to rearrange one resource of one non real time connection. It is noteworthy that we use intuitive definitions for the preemption and the overhead costs. Here, $C_p(x, a)$ is proportional to the number of preempted non real time connections (ζ) and $C_o(x, a)$ accounts solely for the amount of bandwidth rearranged or moved by the preemption. As stated in Section I, the the extra signaling overhead needed to run and manager the preemption operation not only in the link under analysis, but also in others where the extra signaling traffic is transmitted could be used in this analysis as well.

Also, our "feeling" is that more complex costs could still be used. For instance; as pointed in [14], the preemptive priority mechanism is presented in operating systems as eCos, WinCE, VxWorks, QNX, uC/OS, etc., which run on wireless embedded devices; thus, we can use another cost related to the consumption of battery power in the mobile device needed to run the preemption operation. Additionally, the preemption operation activities may also impact of the consumption of battery power require to rearrange the radio resource reallocated among the connections in the mobile device. Hence, another cost can be used to quantify this impact in the cost function.

With $\tau_x(a)$, $p_{xy}(a)$ and $C_x(a)$, using the value iteration algorithm and the uniformization method [24], we can obtain the optimal CAC stationary policy. A stationary policy R,

$$p_{xy}(a) = \begin{cases} \lambda_{ri}\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, i), y = (m_{r1}, \cdots, m_{ri} + 1, \cdots, m_{nr}, e), a = 1; \\ \lambda_{ri}\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, i), y = x, a = 0; \\ \lambda_{ri}\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, i), y = (m_{r1}, \cdots, m_{ri} + 1, \cdots, min(m_{nr}, \theta), e), a = 2; \\ \lambda_{nr}\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, 0), y = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr} + 1, 0), a = 0, \\ & m_{nr} < \lfloor \frac{B - \sum_{i=1}^{N} m_{ri}B_i}{B_{mim}} \rfloor; \\ m_{ri}\mu_x\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, 0), y = (m_{r1}, \cdots, m_{ri} - 1, \cdots, m_{nr}, e), a = 0; \\ m_{nr}\mu_{nrx}\tau_x(a), & x = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr}, 0), y = (m_{r1}, \cdots, m_{ri}, \cdots, m_{nr} - 1, 0), a = 0; \\ 0, & Otherwise. \end{cases} \quad (5)$$

defined by the decision rule $f : \Phi \to A$, prescribes the action $f(x) \in A(x)$ each time the system is observed in the state $x \in \Phi$.

G. Performance Measurement

In this section, we derive the performance measurements used to assess the system performance. The mean carried real time service class i connection traffic is computed as:

$$O_i = \sum_{x \in \Phi; 1 \le e \le N; a=1,2 \in A(x)} (\sum_{i=1}^{N} \lambda_{ri} + \sum_{i=1}^{N} m_{ri}\mu_{ri} + \lambda_{nr} + m_{nr}\mu_{nrx})\pi_x, \quad (10)$$

where π_x ($\forall x \in \Phi$) is the continuous time Markov chain steady state probability distribution under the optimal policy. Giving O_i, we can derive the real time service class i connection blocking probability as follows:

$$P_{bi} = 1 - \frac{O_i}{\lambda_i}. \quad (11)$$

The bandwidth utilization is defined as the ratio between the mean number of occupied channels and the total number of channels, *i.e.*,

$$U = \frac{1}{B} \sum_{x \in \Phi, a \in A(x); m_{ri} > 0 (1 \le i \le N), m_{nr} > 0} (\sum_{i=1}^{N} m_{ri}B_i + b_w(x)m_{nr})\pi_x. \quad (12)$$

The mean number of preempted non real time connections is given by:

$$N_{pd} = \sum_{x \in \Phi; 1 \le e \le N; a=2 \in A(x); \zeta > 0} \zeta\pi_x. \quad (13)$$

III. NUMERICAL RESULTS

In this section, we present numerical results to assess the QoS performance measurements of the proposed optimal CAC developed previously. Here, it is considered a resource sharing system with two service classes: a real time service class and a non real time service class. We present an analysis discussing how variations on the real time blocking costs, preemption cost and overhead cost can impact the system performance. To this end, we vary c_{b1}, c_p, c_o in the range of 1 to 8, but fix $c_p = c_o$. Table I shows the set of the remainder parameters used in the experiments.

Fig.1 shows that the proper selection of c_{b1}, c_p, and c_o greatly affects the optimal CAC performance in such a way

TABELA I
SYSTEM CONFIGURATION.

Parameter	Value
B	10 channels
$1/\mu_{r1}$	10 min
$1/\mu_{nr}$	25 min
$\rho_{r1} = \rho_{nr}$	2
B_1	2 channels
$[B_{min}, B_{max}]$	$[1, 3]$ channels

that for large $c_p = c_o$ and small c_{b1}, higher the real time connection blocking probability and vice-versa. Fig.2 shows that the bandwidth utilization are higher when c_p and c_o are higher than c_o. The reason for this resides in the fact that more ongoing non real time connections are kept in the system for this configuration. It is confirmed in a analysis of the Fig.3, which outlines the mean number of preempted non real time connections versus c_{b1} e $c_p = c_o$. As shown, this QoS performance measurements increases as c_{b1} becomes higher than $c_p = c_o$ and reaches its maximum level when $c_{b1} = 8$ and $c_p = c_o = 1$. It is noteworthy that in this configuration, the optimal CAC has the lowest bandwidth utilization. As shown in Fig.4, the optimal cost increases as the real time blocking cost, preemption cost, and overhead cost increase.
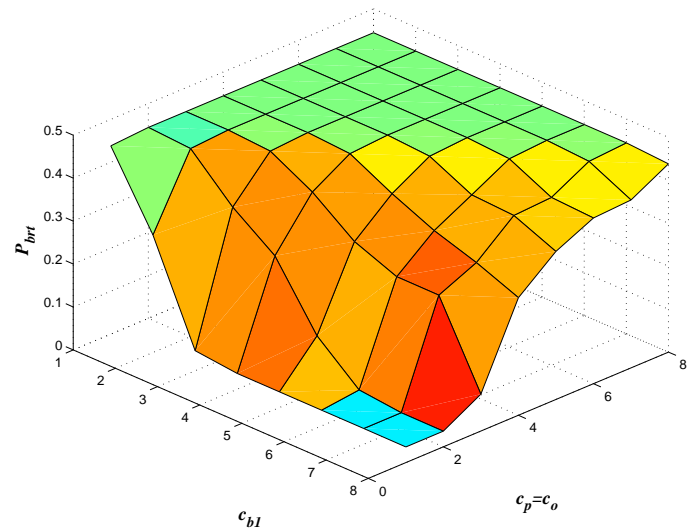

Fig. 1.   Real Time Connection Blocking Probabilities versus c_{b1} e $c_p = c_o$
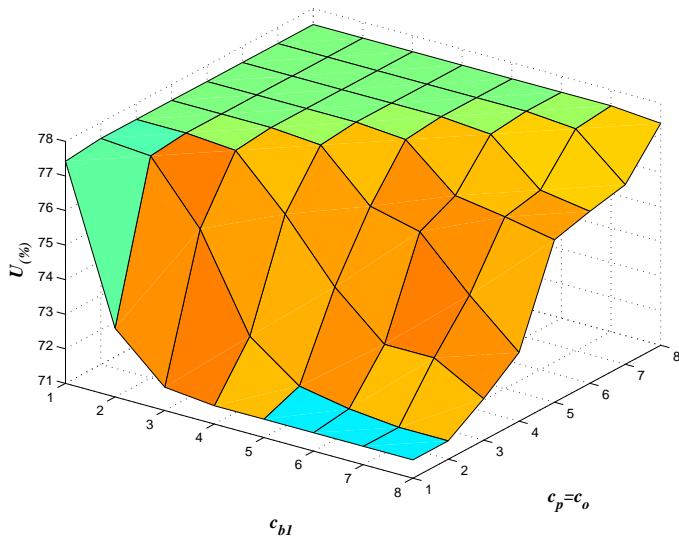
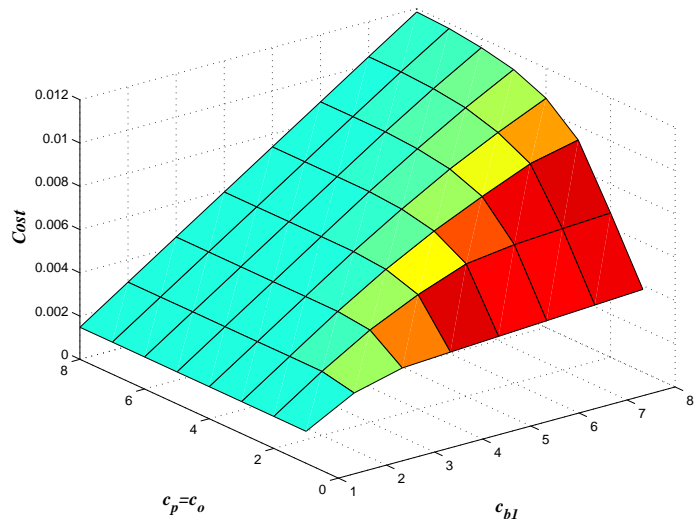Fig. 2.   Bandwidth Utilization versus c_{b1} e $c_p = c_o$

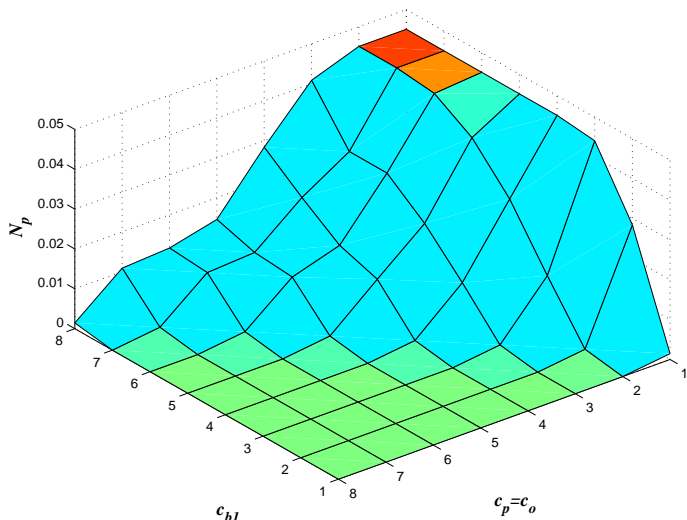Fig. 4.   Optimal Cost versus c_{b1} e $c_p = c_o$

Fig. 3.   Mean number of preempted non real time connections versus c_{b1} e $c_p = c_o$

IV. CONCLUSION

In this paper, we analyzed an optimal CAC in a multi-service resource sharing system like ATM/DSL/WiMax/PLC/Cellular mobile networks. The novelty of our approach is to investigate the impact of preemption operation on CAC designing. This priority is often used to provide resource assurance for high priority service classes and because of that it has been widely employed in many communication and networking systems. However, studies about its effect in CAC decision making has not been carried out so far. Results show that the preemption and overhead costs greatly influences the system performance. It means that when a preemptive priority is taking into account in the design of the CAC, the effect of their set of activities must not be neglected.

REFERÊNCIAS

[1] D. Niyato and E. Hossain, Call Admission Control for QoS Provisioning in 4G Wireless Network: Issues and Approaches, *IEEE Network*, 19, 5, pp. 5–11, 2005.
[2] R. R. M. Comaniciu C, T. V. Lakshman, H. V. Poor, Call Admission Control in Wireless Multimedia Networks, *IEEE Signal Processing Magazine*, 21, 5,51–58, 2004.
[3] K. Mosharaf, I. Lambadaris, J. Talim, A call admission control for service differentiation and fairness management in WDM grooming networks, *Elsevier J. Optical Switching and Networking*,2, 113–126, 2005.
[4] G. L. Choudhury, Analysis of combined voice/data/video operation in cable and DSL access networks: graceful degradation under overload *Performance Evaluation*, v.52, Issues 2-3, pp. 89-103, April 2003.
[5] W. Ni, W. Li, M. Alam, Determination of Optimal Call Admission Control Policy in Wireless Networks, IEEE Transactions on Wireless Communications, v. 8, no. 2, pp. 1038–1044, february 2009.
[6] E. S. Navarro, Y. Lin, V. W.S. Wong, An MDP-Based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks, IEEE Transactions on Vehicular Technology, v. 57, no 2, pp. 1243–1254, March 2008.
[7] V. N. Personé and V. Grassi, Optimal Access Control for Integrated Services Wireless Networks, *Elsevier J. Computer Communications*, 21, 1559-1570, 1998.
[8] S. M. Senouci, A.-L. Beylot, and G. Pujolle, Call Admission Control in Cellular networks: a reinforcement learning solution, *International Journal of Network Management*, 14, 89-103.
[9] F. Yu, V. W.S Wong, V. C.M. Leung, A new QoS provisioning method for adaptive multimedia in cellular wireless networks, *In Proceedings of IEEE INFOCOM*, 2004.
[10] W.-B. Yang and E. Geraniotis, Admission policies for integrated voice and data traffic in CDMA packet radio networks, *IEEE Journal on Selected Areas in Communications*, 12, 4, 654-664, 1994.
[11] S. Singh, V. Krishnamurthy, and H.V.Poor, Integrated voice/data call admission control for wireless DS-CDMA systems, *IEEE Transactions on Signal Processing*, 50, 6, 1483 - 1495, 2002.
[12] B. R. Haverkort, *Performance of Computer Communication Systems: A Model-Based Approach*, John Wiley & Sons, 1999.
[13] B. Li, L. Li, B. Li, K. M. Sivalingam, and X.-R.Cao, Call Admission Control for Voice/Data Integrated Cellular Networks: Performance Analysis and Comparative Study, *IEEE Journal on Selected Areas in Communications*, 4, 22, 706–718, 2004.
[14] S. Tang, and W. Li, An Adaptive Bandwidth Allocation Scheme with Preemptive Priority for Integrated Voice/Data Mobile Networks, *IEEE Transaction on Wireless Communications*, 5, 10, 2006.

[15] W. Song, H. Jiang, W. Zuang, and X. S. Shen, Resource Management for QoS Support in Cellular/WLAN Interworking, *IEEE Network IEEE Network*, 19, 5, 12–18, 2005.

[16] W.Y. Chen, J.L.C. Wu, H.H.Liu, Performance Analysis of Radio Resource Allocation in GSM/GPRS Networks, *In proceedings of IEEE 56th VTC 2002-Fall*, 2002.

[17] X. Fang, D. Ghosal, Performance modeling and QoS evaluation of MAC/RLC layer in GSM/GPRS networks, *In proceedings of ICC03*, 2003.

[18] C. Lindemann and A. Thummler, Performance analysis of the general packet radio service, *Elsevier J. Computer Networks*, 41, 1, 1 - 17, 2003.

[19] W.-Y Chen,J.-L. C. Wu, and L.-L. Lu, Performance Comparisons of Dynamic Resource Allocation With/Without Channel De-Allocation in GSM/GPRS Networks, *IEEE Communication Letters*, 1, 07, 10-12, 2003.

[20] F. A. C-Pérez and L. O.-Guerrero, Capacity Optimization in Wireless Communication Systems With Mixed Platforms, *IEEE Communications Letters*, 4 8, 217-219, 2004.

[21] E. Ng and A. S. Alfa, Optimizing Bandwidth Allocation of Different Traffic Classes for Traffic Between an ISP and a Future Home Area Network, *In proceedings of the Third Advanced International Conference on Telecommunications (AICT07)*, 2007.

[22] E.W.M.Wong, M.Zukerman, Analysis of an optical hybrid switch, *IEEE Communications Letters*, 10, 2, 108 - 110,2006.

[23] H. L. Vu, A.Zalesky, E.W.M.Wong; Z.Rosberg, S.M.H.Bilgrami, M.Zukerman, R.S.Tucker. Scalable performance evaluation of a hybrid optical switch, Journal of Lightwave Technology, 23, 10, 2961 - 2973,2005.

[24] H. C. Tijms, *A First Course in Stochastic Models*, John Wiley & Sons Ltd, 2003.