A Call Admission Protocol for Cellular Networks that Supports Differentiated Fairness

Mokhtar A. Aboelaze

Dept. of Computer Science and Engineering
York University
Toronto, ON. Canada
aboelaze@cs.yorku.ca

Abstract

Call admission protocols play a central role in determining the performance of any network. The call admission protocol must decide either to accept the call or reject it, in the same time it must deals with different classes of calls that require different bandwidth requirements, different quality of service (QoS), and having different priorities. In the same time it must maintain some form of fairness (depends on QoS) and maintain a reasonable utilization of the channel. In this paper, we assume a cellular system and we present a new call admission protocol. Our protocol is simple to implement, and it can support differentiated fairness in call acceptance. We also present a Markov chain representation of a system using our proposed protocol. Finally we present simulation results in order to compare our protocol to previous protocols and show that our protocol can achieve the required differentiated fairness without sacrificing the channel utilization.

1. Introduction

Admission control protocols play a crucial role in the performance of any network. When deciding whether to admit a call or not, many factors must be taken into consideration. Most of these factors are contradictory. A good Call Admission Control (CAC) protocol will try to be fair, fast, reduces customer inconvenience, and produces a good bandwidth utilization leading to increasing revenue for the carrier. The demand for multimedia services in wireless networks has been steadily increasing, so is the research in how to support a certain Quality of Service (QoS) for multimedia applications. The objective here is to limit the number of calls in order to guarantee the requested QoS for each admitted call. Wireless networks share with wire-line networks the need to limit the newly requested calls. However cellular wireless networks suffer from another added complication that is handoff's. When a customer moves from a cell to another cell, we must request a call admission in the new cell the customer moved to, that should be treated differently than a new call request in the cell.

Today's networks to some extent, and definitely's future networks support different types of traffic (data transfer, audio, and video are common in today's networks). Each has its own bandwidth requirements, and its own QoS requirements. Even customers in the same categories, may require different bandwidth and/or different QoS if they are willing to pay more. A good CAC protocol should take these factors into consideration when deciding to admit a call.

Another important factor in call admission is the new calls vs. handoff calls. From a customer point of view it is much less desirable to drop a call in the middle of the connection because of lack of bandwidth in the cell the customer is moving into than to be denied admission at all [5]. Virtually all good CAC protocols gives priority to handoff calls over new calls. Today's networks are moving towards smaller cells in order to increase the capacity and reduces the power.With reducing power and reducing the cell size, calls (customers) are experiencing much more handoff's compared with larger power and larger cells of yesteryears. This is complicated by the fact that handoff calls and new calls may require different bandwidth. A good CAC protocol must give priority to handoff calls over new calls.

The Internet Engineering Task Force (IETF) has proposed two models to guarantee the QoS requested by the users. These two models are called Integrated Service (IntServ), and Differentiated Service (DiffServ).

In the first model (IntServ), network resources are explicitly reserved for each traffic type. Incoming packets at each router are classified according to each type and dealt with accordingly.

In the second model (DiffServ), no explicit reservation is made, instead, the incoming traffic is classified into different set or classes. The network deals with each class according to some priority scheme that is associated with the class type, thus giving a preferential treatment to high priority classes (traffic). Our work in this paper is considered under DiffServ model.

In this paper we present a new CAC protocol for wireless cellular networks. Our protocol takes into consideration the required bandwidth of the call, the available bandwidth into he cell, and the priority level of the requesting customer in deciding whether to admit the call or not. We present a Markov chain to describe the state of the system using our protocol, and we present simulation results to show the performance of our protocol and compare it with other protocols.

The organization of this paper is as follows. In section 2 we present a brief review for previous work in CAC protocols, section 3 is a motivation for our proposed protocol. Section 4 presents our proposed precool and the Markov chain representation of a cellular system using our proposed protocol. Section 5 presents the simulation results. We end our paper with a conclusion and future work.

2. Previous Work

The easiest and most simple admission control is FCFS. In FCFS if a request arrives and there is enough bandwidth to accommodate it, the call is admitted, otherwise it is rejected. FCFS produces a good utilization of the medium, however it has been shown to be biased against calls that require high bandwidth. besides, it does not support priority.

In [14], the authors proposed to divide the bandwidth into segments, and the call requests grouped into different categories,

0-7803-8256-0/04/$20.00 (C) 2004 IEEE

such that a call request in group *i* can only be accepted if there is enough bandwidth in segment *i*. The main problem with this technique is the waste of the bandwidth since we could have unused bandwidth in one segment, and call requests in other segments being turned down.

In [18], a protocol was proposed for admission control in WIND-FLEX [19]. The protocol proposed uses the peak rate of the requested bandwidth, and a sliding window estimate of the current link utilization in order to decide to accept a new call request or not.

The authors in [8] studied the performance of some widely known call admission control protocols under more general (more accurate) assumptions and provided good approximations for the network performance. In [2], the authors proposed a model for heterogeneous multi-class environment that permits call transition between different classes. They also show that under some assumptions, the optimal policy has the shape of *Multi-Priority Threshold Policy*.

In the cutoff priority scheme [16], [12], a portion of the bandwidth is reserved and could be used only for handoff calls. In [17], new calls are admitted with a certain probability that depends on the number of busy channels. The main idea here is to give a priority for handoff calls over new calls.

In another scheme, handoff and new calls are accepted as long as there is enough bandwidth. If there is not enough bandwidth, then either handoff calls are blocked and new calls are pout in a queue [9], or vice versa. Call admission in a *Low Earth Orbit* (LEO) is investigated in [7].

The effect of waiting time in the queue on the new and handoff calls is investigated in [4], they also investigated the effect of the buffer size and the number of guard channels on the system performance for both microcells and macrocells. They proved that good provisioning of the buffering scheme and the number of guard channels can greatly affect the dropoff probability.

Two probability based adaptive algorithms for call admission are presented and analyzed in [21]. While [6], and [15] propose call admission control algorithms taking into consideration the availability of bandwidth in the neighboring cells, thus reducing the call dropoff probability (for handoff calls).

In [10], the authors proposed a distributed algorithm for call admission in which information about the neighboring cells are taken into consideration in admitting any new call. They succeeded in guarantying an upper bound on the call dropping probability and in the same time allowing a high resource utilization.

In [1], the authors investigated the use of a protocol that uses adaptive priority that depends on the link utilization as well as on the traffic characteristic of the application and its tolerance to delay in the network.

A simple but rather efficient algorithm for call admission is presented in [13], where the authors proposed the use of a single buffer to hold the call request if there is not enough bandwidth. The call is held in the buffer until there is enough bandwidth and then admitted, or held in the buffer up to a maximum waiting time then dropped. Their protocol works fine and produces good results if there is not a huge disparity between the requested bandwidths.

3. Motivation

Our proposed protocol is an improvement over the one proposed in [13]. There are two major drawbacks with the protocol presented in [13], the first one is if a high bandwidth demanding call is put in the buffer, then it blocks all other calls until the call is granted (or dropped after long waiting time). In the mean time, many other calls are rejected while there is enough bandwidth to accommodate them. The second drawback is that this protocol aspires for an absolute fairness between all the users, while differentiated service is implemented and required in many networks.

A generally agreed upon measure of fairness is knows as the Jain fairness measure [11] and is expressed as following. For a set of numbers $x_1, x_2, ..., x_N$

$$F = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N \sum_{i=1}^{N} x_i^2} \tag{1}$$

Which lies between *1* and *1/N*. The problem with this definition is that it assumes all x's to have the same weight, i.e. absolute fairness. and is optimal when all x_i*'s* are equal. In order to maintain a differentiated fairness, we have to consider another form of fairness. One possibility is to consider the deviation from a pre-specified set of values for the parameter under consideration and consider only the deviation from these values. For example if the measured values are $x_1, x_2, ..., x_N$, the targeted values are $\gamma_1, \gamma_2, ..., \gamma_N$, then we can consider the fairness over $\hat{x}$, where $\hat{x} = |x_i - \gamma_i|$. Due to the shortage of space in this paper, we will not consider this measure here and will be treated in a subsequent paper.

4. Our Protocol

We assume a cellular system in which the coverage area is divided into cells. There is some overlap between the cells that helps in a smooth handoff. New calls are admitted to each cell when user try to connect and request a specific bandwidth that depends on the application. We assume that the users ask for a specific bandwidth that can not be negotiated. From the user point of view, the call is either admitted, or rejected (busy network). From the network point of view, the user either admitted, rejected, or queued waiting for another user releasing some bandwidth. The queueing period should be small enough such that the user will not notice it.

In a differentiated service model (DiffServ) [3], different users can require different QoS and pay accordingly. Since in this paper we are considering a call admission protocol, we will consider only the quality of service concerned with call admission (call rejection ratio), or call migration (handoff drop ratio).

We assume *N* different classes of customers, each with a different arrival rate, service time, and bandwidth requirements. We

0-7803-8256-0/04/$20.00 (C) 2004 IEEE

also assume that is the arrival rate for a customer in class *i* is Poisson with rate of λ_l customers per second, the service rate of each customer is μ_l seconds, and require a bandwidth B_i. If a call request from class *i* arrives at the base station, if there is enough bandwidth for it and the remaining bandwidth in the cell is more than or equal T_i(a threshold for accepting class i calls), the call, is accepted. Else, if there is not enough bandwidth to accommodate this call, and there are no waiting calls in the buffer, and the remaining bandwidth > T_i, the call is put in the waiting buffer until there is enough bandwidth to be accepted. Else, if there is another call in the buffer, or if the buffer is empty but the remaining bandwidth < T_i, the call is rejected. Thus T_i acts as a parameter to set the priority of class i, the higher the threshold, the less the priority of that class. The priority could be set according to any criterion. It could be set high for customers who are willing to pay more, or could be set high to hadoff calls.

4.1 Protocol description

This protocol is simple and can be described in an algorithmic form as follows:

Assume the capacity of the channel is *C*, class i requires a bandwidth of b_i per call, and its threshold is T_i

```
if ( (remaining >= bj ) && (remaining >= Tj))
  accept the call
else if (remaining < bi) && (remaining >=Ti)
  accept the call and store it in a buffer
  else if (remaining < Tj)
  reject
```

4.2 Markov Chain Representation

A system using the above mentioned protocol, and assuming a Poisson arrival and exponential call time, can be described by a multidimensional Markov chain, *(k+1)* dimensional chain, where *n* is the number of user classes as follows:

The state space consists of (*k+1*)-tuple $(n_1, n_2, \ldots, n_k, n_{k+1})$, such that $\sum_{i=1}^{k} n_i b_i \le C$ where $n_i,\ 1 \le i \le k$ are the number of calls in class *i* admitted in the system, while n_{k+1} is the class of the call waiting in the buffer, if zero, that means the buffer is empty (no calls are waiting). Two constraints on the space state of the markov chain

$$\sum_{i=1}^{k} b_i n_i \le C,\ and \text{ if } n_{k+1} > 0 \Rightarrow b_{n_{k+1}} > C - \sum_{i=1}^{k} b_i n_i > T_{n_{k+1}} \tag{2}$$

The constraint simply states that the bandwidth of the total admitted calls can not be more than the channel bandwidth, and if there is a waiting call, then the remaining bandwidth is greater than or equal to the threshold for the call waiting class (otherwise, it shouldn't have been admitted to the buffer).

The transition probability from state $\boldsymbol{I}=(n_1, n_2, \ldots, n_k, n_{k+1})$ to state $\boldsymbol{I'}=(n'_1, n'_2, \ldots, n'_k, n'_{k+1})$ depends on the system state, the probability can be described as $p(I, I')$, and is shown in the following equation.

$$p(\boldsymbol{I}, \boldsymbol{I'}) = \begin{cases} \mu_j & \{n_i = n'_i \cdot \forall i, i \ne j\} \wedge \{n'_j = n_j - 1\} \wedge \\ & \{\mathrm{n}_{k+1} = n'_{k+1} = 0\} \\ \lambda_j & \{n_i = n'_i, i \ne j\} \wedge \{n'_{k+1} = n_{k+1} = 0\} \wedge \\ & \{n'_j = n_j + 1\} \wedge \left\{ C - \sum_{i=1}^{k} n_i b_i \ge max(b_j, T_j) \right\} \\ \lambda_j & \{n_i = n'_i \forall i\} \wedge \{n'_{k+1} = j\} \wedge \{n_{k+1} = 0\} \\ & \left\{ T_j < C - \sum_{i=1}^{N} n_i b_i < b_j \right\} \\ \mu_j & \{n_i = n'_i\} \wedge \{n'_k = 0\} \wedge \{n_{k+1} = j\} \\ \lambda_j & \{n'_i = n_i \forall ((i,j) \ne l, j)\} \wedge [n'_j = n_j - 1] \wedge \\ & \{n'_l = n_l + 1\} \wedge \{n'_{k+1} = 0\} \wedge \{n_{k+1} = l\} \wedge \\ & \left\{ c - \sum_{i=1}^{n} n_i b_i \ge max(T_l, b_l - b_j) \right\} \\ \mu_j & \{n'_{k+1} = n_{k+1} \ne 0\} \wedge \{n_i = n'_i, i \ne j\} \wedge \\ & \{n'_j = n_j - 1\} \wedge \left\{ c - \sum_{i=1}^{k} n'_i b_i < b_{n_{k+1}} \right\} \end{cases} \tag{3}$$

The first case is an empty buffer, one call is completed and leaves the system. The second case is an empty buffer, the incoming call is admitted right away (there is enough bandwidth, and the remaining bandwidth is more than the threshold). The third case is when the buffer is empty, the incoming call could not be admitted because of lack of bandwidth and will be put in the buffer. The fourth case is when there is a class *j* waiting, and a class *j* completes and leaves the system. The fifth case is when a class *j* customer leaves, and the one waiting in the buffer goes in (the one waiting is a class *l*.) The last case is when a class *j* leaves but the one in the buffer does not go in because there is not enough bandwidth available

5. Simulation Results

We have simulated the above protocol using CSIM [20], we also simulated the protocol in [13] in order to compare our protocol with.

In our simulation, we considered a system with total bandwidth of *C=10* units, two types of customers can arrive, type I needs 2 units of, type II needs 4 units of bandwidth. The service time for class I is 2 time units, and for class II is 4 time units.

Figure 1 shows the rejection ratio vs. arrival rate of class 2 assuming that $C=10$, $b_1=2$, $b_2=4$, $T_1=0$, $T_2=3$ and the arrival rate for class 1 is 2 calls per time. The figure also shows the performance of the protocol proposed in [13]. It is clear from the figure that our protocol gives a better rejection ratio for class 1 (CL1) calls than class 2 (CL2). Note that increasing the arrival

0-7803-8256-0/04/$20.00 (C) 2004 IEEE

rate for CL2 increases its rejection ratio without greatly affecting the rejection ratio of CL1.

Figure 2. shows the utilization vs. the arrival rate of class 2 for the same setting in Figure 1. Note that the utilization is the same as in [13]. which means that most of the bandwidth is being used by the high priority traffic (class I).

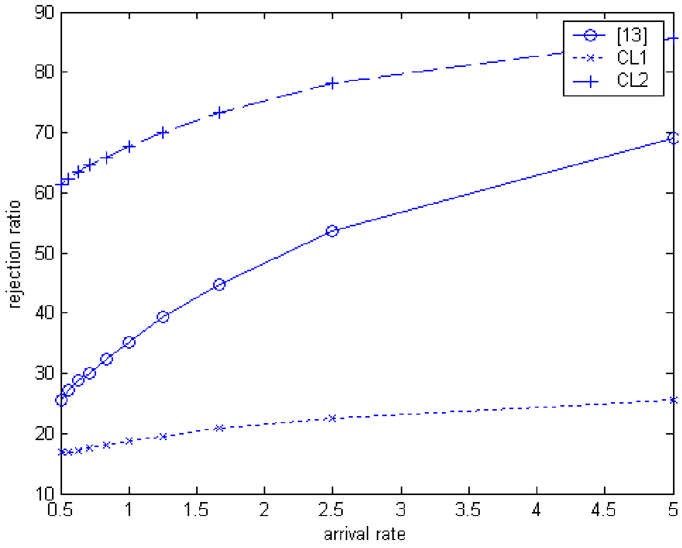

Figure 1. rejection ratio vs. arrival rate

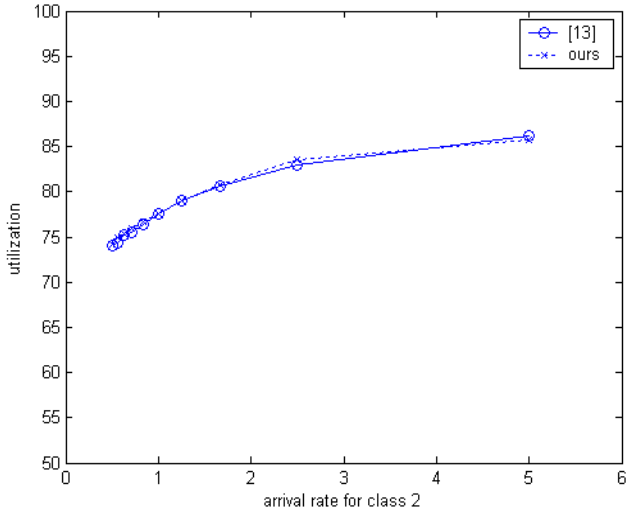

Figure 2. arrival rate vs. utilization for class 2 calls

Figure 3. shows the relation between the threshold for class 2 and the rejection ratio for classes 1 and 2. For this experiment we assumed that the threshold for class 1 is 0, the bandwidth requirements for classes 1 and 2 are 2 and 4 respectively, and the arrival rate for class I is 2 customers per time unit, and for class II is 1 customer per time unit. The bandwidth requirements are 2 and 4 units, and the channel capacity is 10. We can see in Figure 3. that when the threshold for Class II is 0 (same as class I), they have the same rejection ratio, with increasing the threshold for Class II, we notice that the rejection ratio for class I starts to decrease and at the same time the rejection ratio for Class II starts to increase.

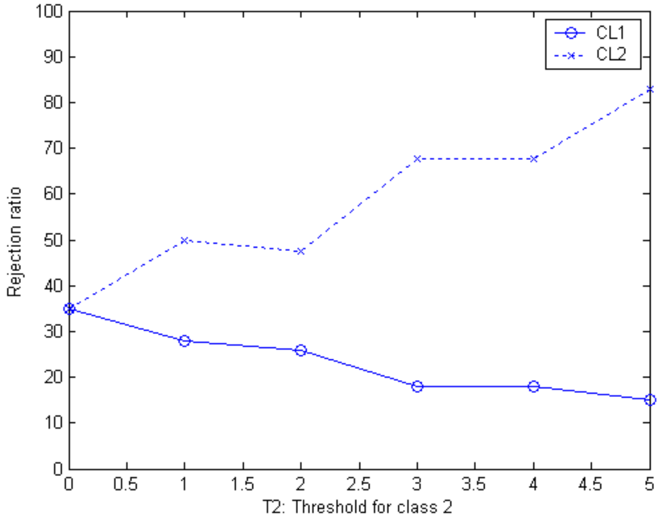

Figure 3. Rejection ratio vs. threshold for class 2

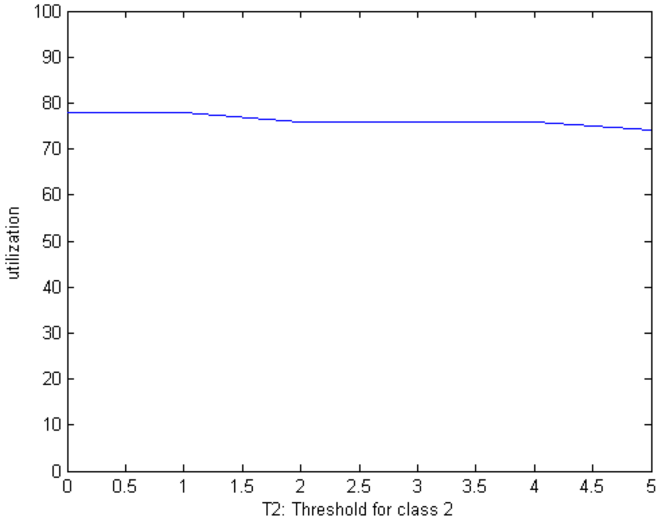

Figure 4. Threshold for Class II vs. Utilization

Figure 4. Shows the threshold for Class II vs. Utilization, here we can notice that the channel utilization changes only slightly with increasing the threshold.

6. Conclusions

In this paper, we presented a new call admission protocol for cellular networks. Our protocol provides a differentiated fairness without sacrificing the channel utilization, we also presented a markov chain representation for a system using our proposed protocol and simulation results to compare our protocol with other call admission protocols.

In this work, we assume no-preemption, and also assumed the bandwidth for a class is constant during the life time of the call and is determined during the admission procedure. For future work we plan to study the effect of pre-emption and the possibility of changing the bandwidth during the life time of the call depending on the channel utilization.

7. References

[1] D. Ayyagari, and A. Ephremides, "Admission Control with Priorities: Approach for Multi-Rate Wireless Systems", *IEEE International Conference on Universal Personal Communication*. Vol 1, pp 301-305 5-9 Oct. 1998.

[2] N. Bartolini, and I. Chamtac "Call Admission Control in Wireless Multimedia Networks" Proceedings of the *13th IEEE International Symposium on Personal Indoor Mobile Radio Communication PIMRC2002*, Vol. 1, pp285-289. Sept. 2002.

[3] S. blake st al, "An Architecture for Differentiated Services", IETF RFC 2475, Dec. 1998

[4] C. Chang, C.-J Chang, and K.-R Lo. "Analysis of a Hierarchical Cellular System with Reneging and Dropping for Waiting New and Handoff Calls" *IEEE Transaction on Vehicular Technology* Vol. 48, No. 4, july 1999.

[5] C. H. Choi, M. I. Kim, and S. J. Kim, "Call Admission Control Using Moving Pattern of Mobile Users for Mobile Multimedia Networks," in *IEEE Conference on Local Computer Networks*, pp 431-440, 2002.

[6] S. Choi, and K. Shin, "Predictive and Adaptive bandwidth Reservation for hand-offs in QoS-Cellular Networks", *Proceedings of ACM SIGCOMM'98* pp 155-166, September 1998.

[7] E. Del Re, R. Fantacci, and G. Giambene, "Handover Queueing Strategies with Dynamic and Fixed Channel Allocation techniques in Low Earth Orbit Satellite Sysems" *IEEE Transaction on Communications* Vol. 74, No. 1 Jan 1999.

[8] Y. Fang, and Y. Zhang "Call Admission Control Scheme and Perfromance Analysis in Wireless Mobile Networks" *IEEE Transaction on Vehicular Technology*, Vol. 51, No. 2, pp 371-382, March 2002

[9] R. A. Guerin "Queueing-Blocking System with Two Arrival Streams and Guard Channels", *IEEE Transaction on Communications* vol 36, No. 2, pp 153-163, 1988.

[10] Y. Iraqi, and R. Boutaba "An Adaptive Distributed Call Admission Control for QoS-Sensitive Wireless Mobile Networks" *Proceedings of the 2000 IEEE Wireless Communication and Networks Conference WCNC*. pp 449-453 Chicago, IL. Sept. 2000

[11] R. jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems" *DEC Research Report* TR-301 Sept. 1984

[12] M. D. kulavaratharasah, and A. H. Aghvami, "Teletraffic Performance Evaluation of Microcellular Personal Communication Networks (PCN's) with Prioritized Handoff Procedures" *IEEE Transaction on Vehicular Technology* vol. 48, No. 1, pp 137-152, 1999

[13] Y. Lai, and Y. Lin, "A Fair Admission Control for Large bandwdith Multimedia Applications", *Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops*, pp 317-322 2nd -5th July, 2002.

[14] Y. C. lai, and S. F. Tsai "A Fair Admission Control for Large bandwidth Multimedia Applications" In *Proceedings of GLOBECOM'2001* pp 317-322, San Antonio, TX Nov. 2002

[15] D. Levine, I. Akyildz, and M. 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 February 1997.

[16] Y. B. Lin, S. Mohan, and A. Noerpel, "Queueing Priority Channel Assignmnet Startegies for Handoff and Initial Access for a PCS Network" *IEEE transactions on Vehicular Technology,* Vol 43, No. 3, pp 704-712, 1994.

[17] R. Ramjee, D. Towsley, and R. Nagarajan, "On Optimal Call Admission Control in Cellular Networks". *Wireless Networks,* vol 3 pp29-41, 1997

[18] G. razzano, and A. Curcio, "Performance Comparison of three Call Admissions Control Algorithms in a Wireless Ad-Hoc Network" *Proceedings of International Conference on Communication Technology ICCT2003*, Vol. 2 pp1332-1336April 9-11 2003.

[19] I. Saarinen, at al "High Bit rate Adaptive (WIND-FLEX) Modem Architcture for Wireless Ad-Hoc Networking in Indoor Environment" *Proceedings of IST Mobile Summit*, Thessaloniki, Greece, June 2002.

[20] H. Schwetman, "CSIM19: A Poweful Tool for Building System Models", *Proceedings of the 2001 Winter Simulation Conference,* 2001

[21] W. Xue-Ping, Z. Jian-Li, Z. Wei, and G.-D Zhang "A Probability-Based Adaptive Algorithm for Call Admission Control in Wireless Network" *Proceedings of the 2003 International Conference on Computer Networks and mobile Computing (ICCNMC'03)* pp197-204, 20-23 Oct. 2003