A Call Admission Protocol for Wireless Cellular Multimedia Networks

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

Fadi A. Aloul
Dept. of Computer Engineering
American University of Sharjah
Sharjah, UAE
faloul@ausharjah.edu

Abstract

In today's wireless networks, different users with different bandwidth (B.W.) requirements, and different quality of service requirements (QoS) are competing for the network bandwidth. The decision to admit or reject a call is made by the call admission protocol. That decision is based on the requested bandwidth, the network utilization, and the priority of the incoming call. In this paper, we assume a cellular system where different users with different B.W. requirements, different QoS requirements and different priorities are competing for admission. We present a new call admission protocol. Our proposed protocol makes the admission decision based on the network utilization, the incoming call priority, and the incoming call's BW requirements. We present simulation results for our protocol and compare it with existing protocols. Our simulation results show that our protocol can better control the different rejection ratio according to priority without sacrificing the network utilization.

1. Introduction

Today's and future generation of wireless networks support different requirements by different users. Although cellular wireless networks started to support ordinary phone calls, today, cellular phones offer many other services such as text messaging, multimedia messages, web browsing, and video transmission. Even for the same type of calls, different customers may have different quality of services requirements based on different cost per user.

Call admission protocol plays a crucial role in the performance of the network. A good call admission protocol must decide to admit or reject new requests. In order to do that, the protocol must take into account many factors, and try to balance the network utilization, and the call rejection ratios, which are usually contradictory. In order to achieve maximum utilization, we must admit any call. In the same time, in order to guarantee a certain call rejection ratio for high priority calls, we must leave some extra bandwidth in the network for high priority calls. Another factor to consider, as the demand for wireless services continues to increase, is that the cell sizes are getting smaller. Small cell size means shorter dwell time and more and more handoffs during the life of the call. That adds another level of complexity since handoff calls should be treated differently than new calls.

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 should 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 on how to support a certain 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.

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 rather than to be denied admission at all [5]. Virtually all CAC protocols give priority to handoff calls over new calls. Today's networks are moving towards smaller cells in order to increase the capacity and reduce the power. With low power and smaller cell size, calls (customers) are experiencing much more handoff's compared with higher power and larger cells of previous years. 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.

In this paper we present a new call admission control protocol for wireless networks that supports priorities and different call rejection ratios for different types of traffic. Our proposed protocol is an extension of the protocol proposed in [1]. Our protocol takes into consideration the priority of the incoming call, and the remaining bandwidth in order to make a decision to accept or reject the call. In our protocol the incoming calls are classified into different categories according to their priority, and the requested bandwidth.

The organization of this paper is as follows. In Section 2, we present a brief overview of previous work in CAC protocols. Section 3 presents our proposed protocol and the Markov chain representation of a cellular system using our proposed protocol. Section 4 presents the simulation setting. It also presents the simulation results of our protocol and compare it with previous protocols. Finally, we conclude in Section 5 and provide some thoughts for future work.

2. Previous Work

In Plain Old Telephone System (POTS), where all calls require the same bandwidth, and all the incoming calls have the same priority, First Come First Serve (FCFS) is used. FCFS produces a good utilization of the communication medium under the above mentioned conditions. However, in cellular networks, where users request different bandwidth depending on the applications, it has been shown that FCFS is biased against calls that require high bandwidth. Furthermore, FCFS does not support priority.

In [10], the authors proposed to divide the bandwidth into segments and group the call requests into different categories, 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 approach is the waste of the bandwidth since we could have unused bandwidth in one segment, while call requests in other segments are rejected.

In [7], the authors proposed a call admission protocol to provide QoS guarantees for multimedia traffic in a heterogeneous Personal Communication System (PCS). network. They assumed two types of traffic: real-time and non-real-time traffic. The bandwidth is divided into channels, each type of traffic may request a different number of channels. The channels are grouped into three groups, one for real-time requests, one for non-real-time requests, and one is combined (for both real-time and non-real-time requests). They also assumed that different cells may have different number of channels. Their protocol depends on varying the boundaries between the three classes of channels in order to satisfy the requested QoS. They also showed that the performance of their protocol could be described by a two-dimensional continuous time Markov chain. They also studied the effect of hot-spots due to the different number of channels in each cell and its effect on new and handoff calls.

The authors of [11] proposed a call admission protocol for integrated voice and data service. They assumed the on/off voice model with silence detection in order to utilize the bandwidth during the off period. Their protocol depends on the use of limited fractional guard channel policy. They also presented a 2-D markov chain model for their protocol.

The effect of admitting a new call on both the uplink and downlink interference is studied in [3]. In their model, they assumed that the recently estimated power distribution is made available to the call admission module. Their protocol depends on estimating the increase in the received as well as transmitted power in both the current cell and the neighboring cells if the call is admitted. Depending on the calculated estimate, the module decide if the increase in the signal to interference power (in case of admitting this call) is acceptable in the current and neighboring cells or not. Based on the results, the module decide to admit/reject the incoming call.

A call admission protocol for wireless networks that support call degradation is introduced in [2]. The protocol admits new calls based on their priority levels and the required bandwidth. The protocol also degrades the on-going calls in order to make bandwidth available for incoming calls. The degradation (both in terms of the amount of bandwidth, and the choice of the calls to be degraded) of the on-going calls also depends on the priority of the on-going calls.

The effect of waiting time in the queue on the new and handoff calls is investigated in [4]. They also explored 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 effect the dropoff probability.

Two probability based adaptive algorithms for call admission are presented and analyzed in [16]. The authors of [6] and [12] proposed call admission control algorithms that take into consideration the availability of bandwidth in the neighboring cells, thus reducing the call dropoff probability (for handoff calls). The work in [14] presents an overview in call admission for DS-CDMA multimedia networks.

In [8], 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 [4], the authors proposed an adaptive distributed call admission control protocol for cellular wireless networks. Their goal is to impose a limit on the call dropping probability regardless of the network load. Their protocol takes into consideration the load in the neighboring cells.

A simple but rather efficient algorithm for call admission is presented in [9], 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 no huge disparity between the requested bandwidths. Our proposed protocol is a variation of this protocol. The modifications results in higher utilization, and introducing priority in the system.

3. Proposed 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 users 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's point of view, the call is either admitted or rejected (busy network). From the network's point of view, the user is either admitted, rejected, or queued waiting for another user to release some bandwidth. If a call request is queued, the queueing time should be small enough for the user not to even notice it. If the queueing time exceeds a maximum time, the call request is dropped and the buffer is cleared. However, as we will see in the simulation, the queueing delay is less than one second in 95% of the cases under the condition described in the simulation section.

We assume N different classes of customers, each with a different arrival rate, service time, and bandwidth requirements. We also assume that the arrival rate for a customer in class i is Poisson with rate of λ_i customers per second. The service time of each customer is exponentially distributed with a mean of $1/\mu_i$ seconds, and require a bandwidth B_i. A buffer of length N is available to store incoming call requests. However, we put a condition on the buffer such that it can not hold more than one call from each class i. If a call from class i arrives, and there is another class i call waiting in the buffer, the arriving call is rejected.

In order to add priority to the system, we assume a threshold T_i for every class of calls i. A call of class i is admitted to the system only if the remaining bandwidth is greater than T_i, thus we can decrease the priority of any class by increasing its threshold. The operation of the protocol is as follows.

Once 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. Otherwise, if there is not enough bandwidth to accommodate this call, and there are no waiting calls in the buffer from class i, and the remaining bandwidth is greater than T_i, the call is put in the waiting buffer until there is enough bandwidth to be accepted. Otherwise, if there is another call of class i in the buffer, or if there are no class i call requests in the buffer, but the remaining bandwidth is less than 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 to handoff calls. The calls in the buffer are granted B.W. according to their priority.

3.1 Protocol Description

This protocol is simple and can be described in an algorithmic form as shown in Figure 1.

```
When a call of class j arrives
if ((remaining >= bj) && (remaining >= Tj))
   accept the call
else if((remaining < bj) &&(remaining >= Tj)
&& (no requests of class j in buffer))
   accept the call and store it in a buffer
else
   reject

When a call is completed
Check the calls in the buffer in descending
level of priority for a waiting class j call
if ((remaining >= bj)
   accept the call
```

Figure 1. Protocol Description. Assume the capacity of the channel is C, class i requires a bandwidth of b_i per call, and its threshold is T_i. remaining **is the remaining bandwidth in the cell.**

3.2 Markov Chain Representation

A system using the above mentioned protocol, and assuming a Poisson arrival and exponential call time, can be described by a $2N$-dimensional Markov chain, where N is the number of different categories of traffic.

We assume that there are N different classes. Without loss of generality, we assume that class 1 has the highest priority and classes are arranged in descending order of priority (class N is the lowest priority class). Each class requires a bandwidth of b_i, and has a threshold of T_i. The arrival rate of class i is λ_i and the average service time of class i is $S_i = 1/\mu_i$.

The state space consists of ($2N$)-tuple $\langle n_1, n_2, \ldots, n_N, m_1, m_2, \ldots, n_N \rangle$, such that

$$\sum_{i=1}^{M} n_i b_i \leq C \tag{1}$$

where n_i $(1 \leq i \leq N)$ is the number of calls of class i admitted in the system, while $m_i \in \{0, 1\}$ is equal to 1 if there is a waiting request of class i in the buffer, otherwise 0. Then,

$$P(\Gamma_1, \Gamma_2) = \begin{cases} \mu_i & \begin{array}{c} [(n'_i = n_i - 1) \wedge (n'_j = n_j) \wedge (m'_k = m_k) \wedge \forall k((m_k = 0) \vee (R < b_k))] \vee \\ [(n'_{j'} = n_{j'}, j' \neq l, i) \wedge (n_l = n_l + 1) \wedge (n'_i = n_i - 1) \wedge \\ (m'_l = m_l - 1) \wedge \forall k' > l, ((m_{k'} = 0) \vee (R < b_{k'}))] \end{array} \\ \lambda_i & \begin{array}{c} [(m'_k = m_k) \wedge (n'_i = n_i + 1) \wedge (R > K_i)] \vee \\ [(n'_k = n_k) \wedge (m'_i = m_i + 1) \wedge (T_i \leq R < b_i)] \vee \\ [(n'_k = n_k) \wedge (m'_k = m_k) \wedge (R < T_i)] \end{array} \end{cases} \tag{2}$$

Where $P(\Gamma_1, \Gamma_2)$ is the transition probability between two states Γ_1 and Γ_2 such that

$$\Gamma_1 = (n_1, n_2, \ldots n_N, m_1, m_2, \ldots, m_N) \text{ and} \quad (3)$$

$$\Gamma_2 = (n'_1, n'_2, \ldots n'_N, m'_1, m'_2, \ldots, m'_N) \quad (4)$$

Where the subscript j implies $\forall j, 1 \le j \le N, j \ne i$, k implies $\forall k, 1 \le k \le N$, R is the remaining bandwidth, and $K_i = max(T_i, b_i)$.

The transition probability can be described as follows: the top conditions state the two possible outcomes when a call is ended, while the bottom three conditions describe the three different outcomes when a call arrives.

When a call is ended; either nothing happens, that is the case if there are no waiting calls, or when the waiting calls require bandwidth more than what is available. Or, one of the waiting calls in the buffer is accepted. When a call arrives, there are three different outcomes: (i) It will be accepted if there is enough bandwidth and the remaining bandwidth is greater than the call threshold. (ii) It will be out in a buffer if there is not enough bandwidth, but the remaining bandwidth is greater than the call threshold. (iii) The call will be rejected if the remaining bandwidth is less than the call threshold.

4. Simulation Setting

We have simulated the above protocol using CSIM [15], we also simulated the protocol in [9] in order to compare our protocol with.

In our simulation, we considered a system with total bandwidth of C = 15Mbps. We simulated two different types of traffic [13]. First, phone calls with bandwidth requirements of 30 Kbps, and a call duration of 3 minutes (180 seconds). Second, data transmission with bandwidth requirements of 20Kbps and an average duration of 30 seconds. For each of these two types, we considered new calls and handoff calls, for a total of 4 different types of traffic. Also, in choosing the different parameters for the system to work under, we assumed that at equilibrium, 75% of the bandwidth is consumed by voice calls, and 25% of the bandwidth is consumed by data connections.

4.1 Waiting time in the buffer

We studied the effect of waiting time in the buffer. In any practical system, and especially for voice calls, users can not wait for a long time until the call is admitted to the system. After some waiting time in the buffer, the call will be rejected and a busy signal is returned to the caller. However, in our simulation, and with utilization going up to 99%, and rejection ratio up to 20%, the waiting time in the buffer was negligible (less than 1 second for 99% of the calls). That makes the assumption in the analytical model to be valid, and the model represent the actual system to a reasonable degree of accuracy.

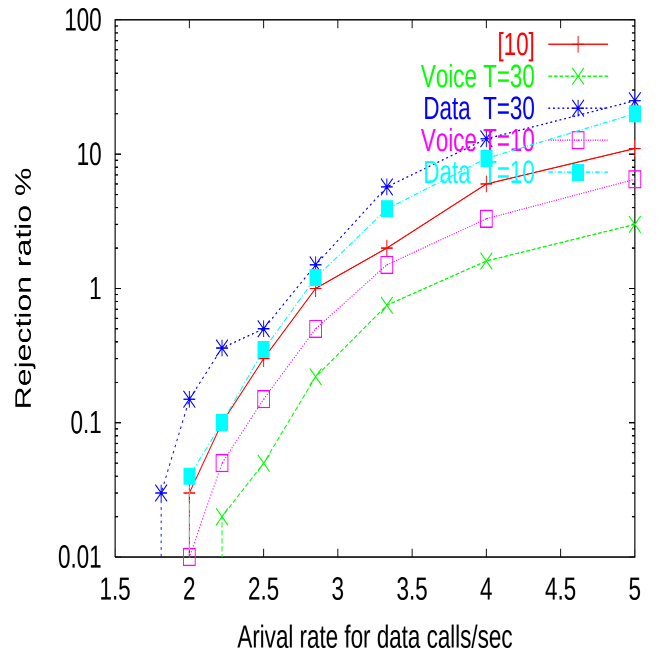

Figure 2. Rejection Ratio vs. Arrival rate for handoff voice calls

4.2 Simulation Results and Discussion

Figure 2 shows the relation between arrival rate of data traffic and rejection ratio for the different classes of traffic, and compares it with the result from [9]. We did not assign different priorities to new and handoff calls; thus there are practically two different types of traffic, voice and data. The arrival rate for both new and handoff voice calls is kept fixed at 1 call per second. The arrival rate for both the new and handoff data calls varies from 1.5 to 5 calls per second. The graph labeled [9] is the performance of the protocol in [9]. In [9] the protocol can not support priority among the different classes, the rejection ratio is the same for the two types. Thus, increasing the arrival rate of data traffic increases the rejection ratio for both data calls as well as voice calls by the same ratio. In our protocol, we assigned different priorities for voice and data calls. The voice calls are assigned a higher priority than data calls by setting the threshold for voice calls to 0, while the threshold of data calls is set at 10Kbps and 30Kbps (if the remaining bandwidth is less than 10/30Kbps, requests for data calls are turned down). We can see in Figure 2 that the increase of data traffic affects the rejection ratio for data calls much more than it affect the voice calls. By changing the threshold value we can control (isolate) the effect of increasing a low priority traffic on a high priority traffic. For example when the threshold for data calls is set to 30Kbps, there is a factor of 10 between the rejection ration for voice calls and data calls. By adjusting the threshold, we can have different rejection rations for different traffic types. Finally, we would like to mention that the utilization of the bandwidth in this experiment ranged from 82% to 99.5%.

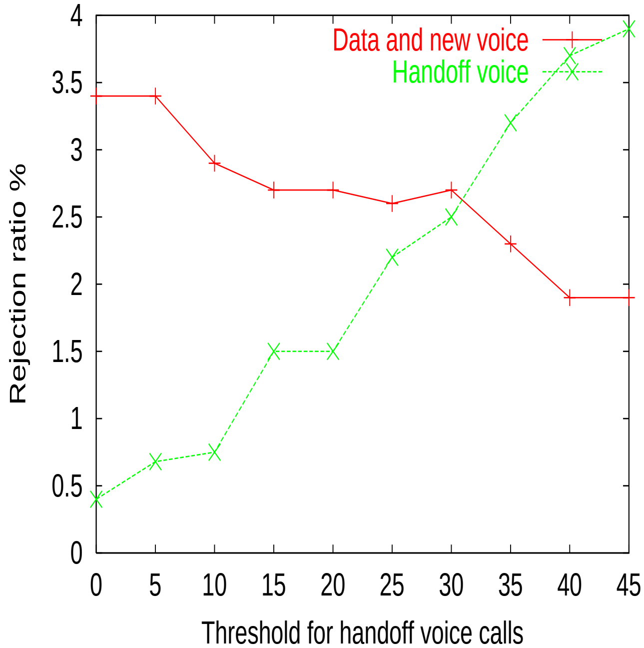

Figure 3. Rejection ratio vs. Threshold for handoff voice calls

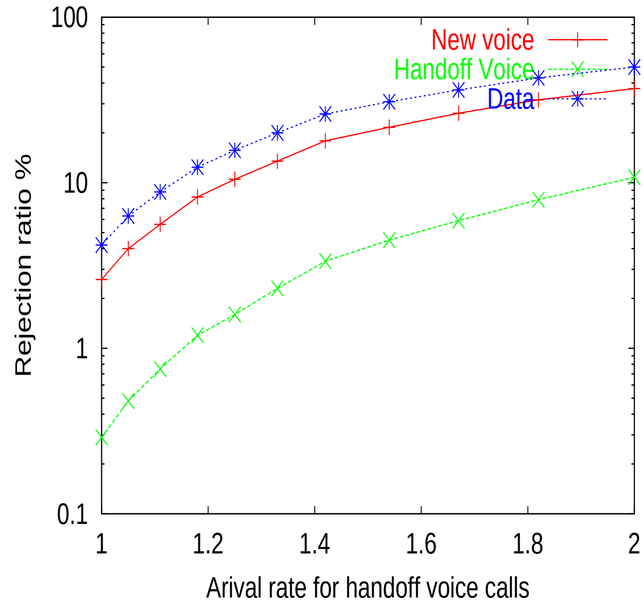

Figure 4. Rejection ratio vs. arrival rate for handoff calls

In Figure 3, we set the new voice calls rate to 1 call per second, and the data calls to 3.33 calls per second (in order to maintain the 75/25% ratio between voice calls and data calls). The new handoff calls rate is set to 1.5 calls per second (that is 50% higher than the new call arrival rate for voice calls).

The threshold for new voice calls is set to 40Kbps, while the threshold for data calls was set to 20Kbps (that is set to have the same priority for new voice and data calls). The threshold for handoff voice calls varies from 0-45Kbps. For a threshold of 0, we notice that the rejection ratio for handoff calls is less than 0.5%, while it is almost 3.5% for new calls and data calls. By increasing the threshold for handoff calls, the rejection ratio for handoff calls starts to increase and that of the new voice and data calls starts to decrease (they are equal at almost 30 Kbps). As the threshold for handoff class continue to increase, the rejection ratio of new and data calls becomes less than that of the handoff calls (of course this is a situation that we would like to avoid in real-life).

Figure 4 shows the relation between arrival rate for handoff calls vs. the rejection ratio for data and new voice calls. In this case, arrival rate for new voice calls is set to 1 call/sec, while that of the new and handoff data calls is set to 3.33 calls/sec.

Handoff calls varies from 1 to 2 calls per second (that is a maximum of 100% increase for handoff calls over new calls). The threshold for handoff calls is set to 0, threshold for new voice calls is set to 30Kbps, while that of new/handoff data calls is set to 40Kbps.

By increasing the arrival rate for handoff calls that result of increasing the rejection ratio for all types of calls, but not with the same rate. The rejection ratio for handoff voice calls is kept below that of any other traffic type, thus giving the handoff voice calls the highest priority among the four types of traffic. By changing the different thresholds, we can control the different priority levels assigned to any traffic types.

Next, we consider a case with a wide variety of bandwidth requirements. We use the combination of traffic in [2] but without the video on demand. The reason for excluding the video on demand is because It has the same bandwidth requirements as the file transfer, and is not suitable for roaming cellular devices.

The mix we simulated is: voice, video conferencing, e-mail/fax, data, and file transfer. The average call duration is 3, 5, 0.5, 3, and 2 minutes, respectively. The BW requirements are (min. requirements) 30, 256, 10, 64, and 1000Kbps, respectively.

In our simulation, we choose the arrival rate as follows: The total load on the network is in the range of 95%. 50% of the load is coming from voice calls, the rest of the bandwidth is equally loaded by the four other types of traffic. The reason for that choice is we believe that the voice calls still represent the majority of cellular networks load.

Figure 5 shows the result of our simulation. The protocol in [9] produces an equal rejection ratio for all the different types of traffic. We fixed the threshold for FTP to 0, and varied the threshold for the other types of traffic (same threshold for all other types of traffic). We can control the rejection ratio for a very high BW demanding application like FTP. We plot only

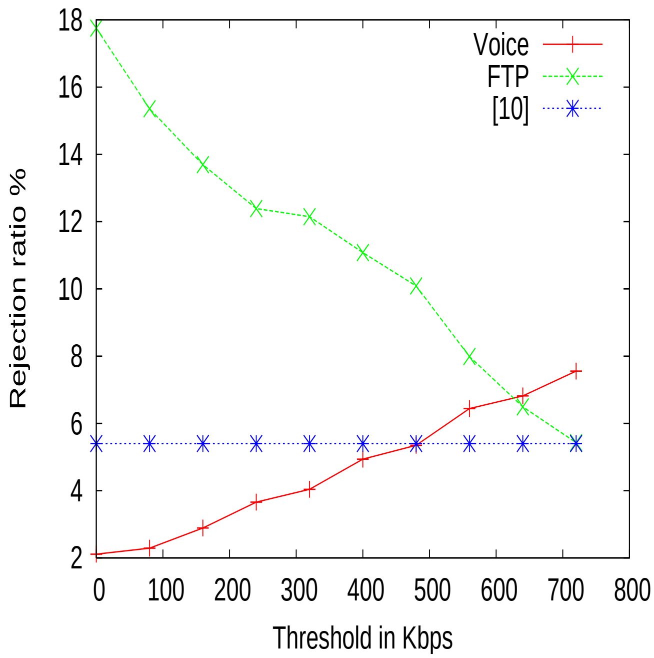

Figure 5. Rejection ratio vs. threshold

the rejection ratio for voice calls, however the rejection ratio for all other types of traffic is within 5% of the voice call. FInally, we should note that the utilization of the network ranged from 88-93% during the simulation.

5. Conclusions

In this paper, we presented a new call admission protocol for cellular networks. Our protocol provides different call admission/rejection ratios to different types of traffic according to their priority 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 assume 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 lifetime of the call depending on the channel utilization.

6. References

[1] M. Aboelaze, "A Call Admission Protocol for cellular networks that supports differentiated fairness," in *Proc. of the IEEE Semi-annual Vehicular technology Conference* (IEEEVTC), Milan. Italy, June 2004.

[2] M. Aboelaze, A. Elnaggar, and M. Musleh, "A priority based call admission control protocol with call degradation for cellular networks," in *Proc. of The International Symposium on Wireless Communication Systems* (ISWCS), Port Luis, Mauritius, 2004.

[3] S. Aissa, J. Kuri, and P. Mermelstein, "Call admission on the uplink and downlink of a CDMA system based on total received and transmitted powers," in *IEEE Transactions on Wireless Communication*. 3(6), 2407-2416, November 2004.

[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," in *IEEE Transactions on Vehicular Technology*, 48(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 *Proc. of the IEEE Conference on Local Computer Networks*, 431-440, 2002.

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

[7] Y.-R. Huang and J.-M. Ho, "Distributed call admission control for a heterogeneous PCS network," in *IEEE Transactions on Computers*, 51(12), 1400-1409, December 2002.

[8] Y. Iraqi and R. Boutaba "An Adaptive Distributed Call Admission Control for QoS-Sensitive Wireless Mobile Networks," in *Proc. of the IEEE Wireless Communication and Networks Conference* (WCNC), 449-453, September 2000.

[9] Y. Lai and Y. Lin, "A Fair Admission Control for Large bandwidth Multimedia Applications," in *Proc. of the 22^{nd} International Conference on Distributed Computing Systems Workshops*, 317-322, July 2002.

[10] Y. C. Lai and S. F. Tsai "A Fair Admission Control for Large bandwidth Multimedia Applications," in *Proc. of GLOBECOM*, 317-322, November 2002.

[11] C. W. Leong, W. Zhuang, Y. Cheng, and L. Wang, "Call admission control for integrated on/off voice and best-effort data services in mobile cellular communication," in *IEEE Transactions on Communication*, 52(4), 778-790, May 2004.

[12] D. Levine, I. Akyildz, and M. Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks Using the Shadow Cluster Concept", in *IEEE/ACM Transactions on Networking*, 5(1), February 1997.

[13] C. Oliveira, J. Kim, and T. Suda "An Adaptive Bandwidth Reservation Scheme for High-Speed Multimedia Wireless Networks", in *IEEE Journal on Selected Areas in Communication*, 16(6), 858-874, August 1998.

[14] R. Rao, C. Comaniciu, T. Lakshman, and H. Poor "Call Admission Cocontrol in wireless multimedia networks" in *IEEE Signal Processing Magazine*, 51-58, September 2004.

[15] H. Schwetman, "CSIM19: A Powerful Tool for Building System Models," in *Proc. of the Winter Simulation Conference*, 2001.

[16] W. Xue-Ping, Z. Jian-Li, Z. Wei, and G.-D Zhang "A Probability-Based Adaptive Algorithm for Call Admission Control in Wireless Network," in *Proc. of the International Conference on Computer Networks and Mobile Computing* (ICCNMC), 197-204, October 2003.