Performance Evaluation of a Distributed Call Admission Control for QoS-Sensitive Wireless Mobile Networks

Youssef Iraqi and Raouf Boutaba
University of Waterloo, Canada
{iraqi,rboutaba}@bbcr.uwaterloo.ca

Abstract

This paper presents a performance evaluation of a distributed call admission control framework developed for cellular mobile networks. The call admission control algorithm involves not only the original cell (handling the new admission request) but also a cluster of neighboring cells. The neighboring cells provide significant information about their ability to support the new mobile user in the future. This distributed process allows the original cell to make a more clear-sighted admission decision for the new user. Simulations are provided to show the improvements obtained using our framework.

1 Introduction

Cellular mobile networks have to continue supporting their mobile users after they leave their original cells. This rises a new challenge to Call Admission Control (CAC) algorithms. A call admission process should not only take into consideration the available resources in the original cell but also in neighboring cells as well.

Mobile users are in a growing demand for multimedia applications, and the next generation wireless networks are designed to support such bandwidth greedy applications. The (wireless) bandwidth allocated to a user will not be fixed for the lifetime of the connection as in traditional cellular networks, rather the base station will allocate bandwidth dynamically to users. Many evolving standards for Wireless Broadband Systems, UMTS and IMT2000 have proposed solutions to support such capability.

Call admission schemes that involve other cells in the decision process can be divided into two categories. Implicit involvement schemes and explicit involvement schemes. In the implicit involvement schemes, cells are involved implicitly in the CAC process. An example of such scheme is the well known Guard Channel scheme [4, 3] where each cell reserves a number of channels (or bandwidth units) for exclusive use by handoff users. In this scheme cells are implicitly involved in the CAC process. In the explicit category, cells are explicitly involved in the CAC process by exchanging important information about users and about their capabilities to support these users. An example of such a scheme is presented in [5].

Several call admission algorithms have been proposed for wireless networks to support multimedia users with dynamic bandwidth requirements (e.g. [2]). These algorithms take only local information in the admission decision process, which result in a high call dropping probability. To reduce the call dropping probability, few other CAC algorithms which take into consideration information from neighboring cells have been proposed [7][1][8]. However, those algorithms only support users with fixed bandwidth requirements.

In [5] we have proposed a Distributed Call Admission Control scheme designed for wireless mobile multimedia networks with dynamic bandwidth allocation. The call admission process involves the cell that receives the call admission request and a cluster of neighboring cells so the user will not be dropped due to handoffs. Consequently, the network will provide a low call dropping probability while maintaining a high resource utilization.

In this paper, we present a performance evaluation of the CAC scheme in terms of Average Bandwidth Utilization (ABU) and Call Dropping Probability (CDP). Other parameters such as Call Blocking Probability (CBP) are also investigated. We compare our scheme with the Guard Channel scheme [4, 3]. We demonstrate that for multimedia support, explicit involvement of cells gives better performance in terms of CDP, CBP and ABU.

The paper is organized as follows. In section 2, we describe the model of the system considered in this paper. In section 3 we present the overall call admission control scheme involving a cluster of neighboring cells. Section 4 discusses the conducted simulation parameters and presents a detailed analysis of the obtained results. Finally, section 5 concludes the paper.

1530-1346/01 $10.00 © 2001 IEEE

2 System Model

We consider a wireless/mobile network with a cellular infrastructure that can support mobile terminals running applications which demand a wide range of resources. Users can freely roam the network and experience a large number of handoffs during a typical connection. We assume that users have a dynamic bandwidth requirement. The wireless network must provide the requested level of service even if the user moves to an adjacent cell. A handoff could fail due to insufficient bandwidth in the new cell, and in such case, the connection is dropped.

To reduce the call dropping probability, we proposed in [5] to make neighboring cells participate in the decision of a new user admission. Each involved cell will give its local decision and finally the cell where the request was issued will decide if the new request is accepted or not. By doing so, the new admitted connection will have more chances to survive after experiencing handoffs.

We use the notion of a cluster similar to the shadow cluster concept [6]. The idea is that every connection exerts an influence upon neighboring base stations. As the mobile terminal travels to other cells, the region of influence also moves. The set of cells influenced by a connection are said to constitute a cluster. Each user in the network, with an active connection has a cluster associated to it. The cells in the cluster are chosen by the cell where the user resides. The shape and the number of cells of a user's cluster depend on factors such as user's current call holding time, user's QoS, terminal trajectory and velocity.

We consider a wireless network where the time is divided in equal intervals at $t = t_1, t_2, ..., t_m$. Let j denote a base station in the network, and x a mobile terminal with an active wireless connection. Let $K(x)$ denote the set of cells that form the cluster for the active mobile terminal x. We denote $P_{x,j,k}(t) = [P_{x,j,k}(t_0), P_{x,j,k}(t_1), ..., P_{x,j,k}(t_{m_x})]$ the probability that mobile terminal x, currently in cell j, to be active in cell k, and therefore under the control of base station k, at times $t_0, t_1, t_2, ..., t_{m_x}$. $P_{x,j,k}(t)$ represents the projected probabilities that a mobile terminal will remain active in the future and at a particular location. It is referred to as the Dynamic Mobile Probability (DMP) in the following. The parameter m_x represents how far in the future the predicted probabilities are computed.

Those probabilities may be function of several parameters such as: residence time of mobile x in cell j, handoff probability, cell size and user mobility profile.

For each user x in the network, the cell that is responsible for, decides the size of the cluster $K(x)$, those are the cells involved in the CAC process, and sends the DMPs to all members in $K(x)$. The cell must specify if the user is a new one (in which case the cell is waiting for responses from the members of $K(x)$) or not.

3 Call Admission Control

3.1 Local Call Admission Control

User's traffic can be either voice, data or video. At each time t_0 each cell, in a cluster $K(x)$ involved in our CAC process for user x, makes a local CAC decision for different times in the future $(t_0, t_1, ..., t_{m_x})$. Based on these CAC decisions, we call Elementary Responses, the cell makes a final decision which represents its local response to the admission of user x in the network. The cell can apply any local call admission algorithm to compute the elementary responses.

We denote $r_k(x,t)$ the elementary response of cell k for user x for time t. If for user x, cell k has a response $r_k(x,t)$ for each t from t_0 to t_m with a corresponding DMPs $P_{x,j,k}(t_0)$ to $P_{x,j,k}(t_m)$, then to compute the *final response* those *elementary responses* are weighted with the corresponding DMPs. The *final response* from cell k to cell j concerning user x is then :

$$R_k(x) = \frac{\sum_{t=t_0}^{t=t_{m_x}} r_k(x,t) \times P_{x,j,k}(t) \times C_k(x,t)}{\sum_{t=t_0}^{t=t_{m_x}} P_{x,j,k}(t)} \quad (1)$$

where $C_k(x,t)$ is the confidence that has cell k about the elementary response $r_k(x,t)$. Cell k, then, sends the response $R_k(x)$ to the corresponding cell j. A detailed description of the algorithm can be found in [5].

3.2 Distributed Call Admission Control

Here the decision takes into consideration the responses from all the cells in the user cluster. The admission process concerns only new users seeking admission to the network and not already accepted users.

We assume that cell j has already decided the cluster $K(x)$ and that cell j has already assigned to each cell k in the cluster $K(x)$ a weight $W_k(x)$. Each weight represents the importance of the contribution of the associated cell to the global decision process. We suggest to use the following formula to compute the weights $W_k(x)$ [5]:

$$W_k(x) = \frac{\sum_{t=t_0}^{t=t_{m_x}} P_{x,j,k}(t)}{\sum_{k' \in K} \sum_{t=t_0}^{t=t_{m_x}} P_{x,j,k'}(t)} \quad (2)$$

The *final decision* of the call admission process for user x is based on

$$D(x) = \sum_{k \in K} R_k(x) \times W_k(x) \quad (3)$$

If $D(x)$ is higher than a certain threshold, we call acceptance threshold (AT), then, user x is accepted; otherwise the user is rejected.

4 Performance Evaluation

We compare our algorithm (DCAC) with the guard channel (GC) scheme where each cell reserves a number of channels for exclusive use by handoff users. We have conducted many simulations to evaluate the performance of the DCAC in different situations.

We simulate a system that uses the GC scheme. We change the number of reserved channels (from 0% to 100% step 2%) for each simulation and we compute several important parameters like the Call Dropping Percentage, the Call Blocking Percentage and the Average Bandwidth Utilization. Also we simulate a system that uses our Distributed Call Admission Control scheme, we change the value of the acceptance threshold (from 0.4 to 0.7 step 0.01) for each simulation and we compute the same parameters.

4.1 Simulation parameters

For simplicity, we evaluate the performance of the two schemes for mobile terminals which are traveling along a highway. This is a simple environment representing a one-dimensional cellular system. In our simulation study we have the following simulation parameters and assumptions:

1. The time is quantized in intervals $\tau = 10s$
2. The whole cellular system is composed of 10 linearly-arranged cells, laid at 1-km intervals. Cells are numbered from 1 to 10. Cell 1 and 10 are connected so that the whole cellular system forms a ring architecture. This allows to avoid the uneven traffic load that will be experienced by cell 1 and 10 otherwise.
3. Each cell has a fixed capacity of 50 channels.
4. Connection requests are generated in each cell according to a Poisson process with rate λ (connections/second). A newly generated mobile terminal can appear anywhere in the cell with equal probability.
5. Mobile terminals can have speeds of: 70, 90, or 105 km/h. The probability of each speed is 1/3, and mobile terminals can travel in either of two directions with equal probability.
6. We consider three possible types of traffic: voice, data, and video. The probabilities associated with these types are R_{voice}, R_{data} and R_{video} respectively. The number of bandwidth units (BUs) required by each type is: $B_{voice} = 1$, $B_{data} = 5$, $B_{video} = 10$.
7. Connection lifetimes are exponentially-distributed with a mean value equal to 180 seconds.

For the Distributed Call Admission Control scheme we assume also that:

1. $m_x = 18$. This means that the DMPs are computed for 18 steps in the future.
2. The size of the cluster $K(x) = 2$. This means that one cell in the direction of the user along with the cell where the user resides form the cluster.
3. The DMPs are computed as in [7].
4. All users have the same acceptance threshold.

The offered load per cell, L, is defined as connection generation rate × connections' bandwidth × average connection lifetime, i.e.,
$L = (R_{voice}.B_{voice} + R_{data}.B_{data} + R_{video}.B_{video}).\lambda.180$.

Knowing the average connection lifetime and bandwidth, we choose the connection generation rate to have a cell load of 100% and we simulate a total of 24 hours of real-time highway traffic for each simulation.

4.2 Simulation results

4.2.1 First set of simulations: data users only

In this first set of simulations we assume that only one class of service is used by all users. We compare the performance of DCAC and GC with only data users, which means that the probabilities associated with the three types of traffic are $R_{voice} = 0$, $R_{data} = 1$ and $R_{video} = 0$ respectively.

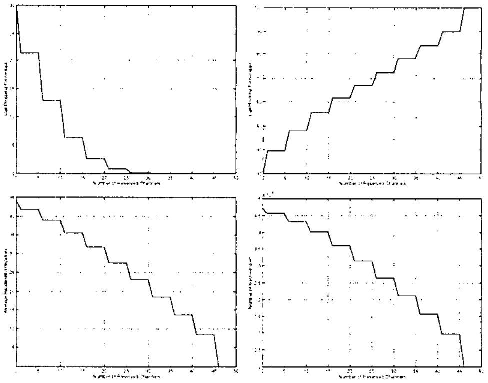

Figure 1. (1.a, 1.b, 1.c and 1.d) Simulation results for GC according to the number of reserved channels

Figure 1 depicts the performance results of the GC scheme. Figure 1.a (top left curve) presents the Call Dropping Percentage experienced by data users according to the number of reserved channels. As expected, when the number of reserved channels increases the CDP decreases. We can notice that since the number of bandwidth units used by data users is 5, the guard channel gives stepped curves with a step of 5. As indicated by figure 1.a some CDP values can not be obtained using the GC scheme. For example, using this scheme, and under the considered cell load, we can not have a CDP of 10%.

Figure 1.b (top right curve) depicts the Call Blocking Percentage experienced by data users according to the number of reserved channels. Of course, when the number of reserved channels increases the CBP increases also which affects the average number of resources used as depicted by figure 1.c (bottom left curve).

Figure 1.d presents the number of finished data users which is the number of data users that were able to access the network and to finish their calls without being dropped. The number of finished data users decreases as the number of reserved channels increases.

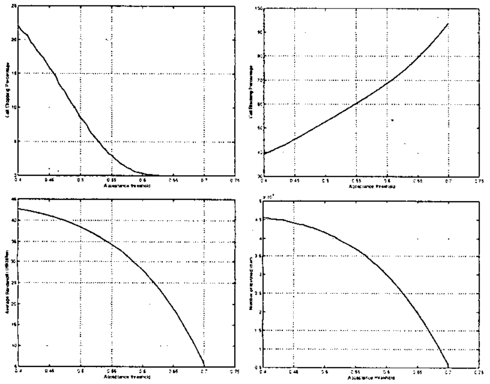

Figure 2. (2.a, 2.b, 2.c and 2.d) Simulation results for DCAC according to the acceptance threshold

Figure 2 depicts the performance results of the DCAC scheme. In this scheme only users that have a decision value (see eq. 3) above the acceptance threshold (AT) are admitted to the network.

Figure 2.a presents the Call Dropping Percentage experienced by data users according to the value of AT. As expected, when AT increases, a smaller number of users are admitted to the network and hence users incur lower CDP. At the contrary of the GC scheme, the DCAC allow for any value of the CDP. Indeed, by choosing the right AT value, we can obtain any desired (and possible) CDP value.

Figure 2.b depicts the Call Blocking Percentage experienced by data users according to the value of AT. Of course, when the value of AT increases the CBP increases also. This is because we require a higher value of the decision results. Figure 2.c shows as expected the decrease in the average bandwidth utilization incurred when the value of the acceptance threshold increases.

Figure 2.d presents the number of finished data users which is the number of data users that were able to access the network and to finish their calls without being dropped. tradeoff

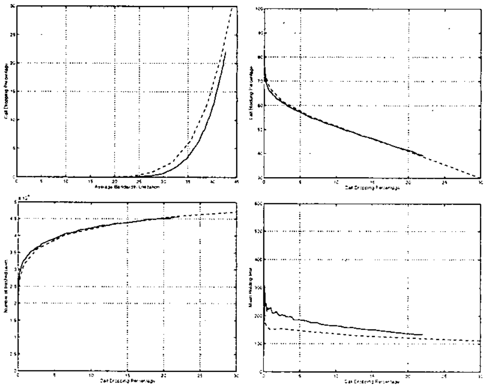

Figure 3. (3.a, 3.b, 3.c and 3.d) Comparing the two schemes (data traffic only)

Figure 3 presents the comparison results of the two schemes. Solid lines refer to the DCAC scheme and dotted lines refer to the GC scheme.

Figure 3.a depicts the Call Dropping Percentage experienced by data users according to the average bandwidth utilization achieved by the two schemes. As presented by figure 3.a, for the same achieved average bandwidth utilization the DCAC scheme allows for a lower CDP. In other words, the DCAC is able to achieve the same CDP while allowing a higher resource utilization. It is important to notice that there is always a tradeoff between the CDP, CBP and average resource utilization. So, we can not have the lower CDP and at the same time the higher resource utilization and a lower CBP. A good scheme is the one that optimizes one parameter without sacrificing the other parameters.

Figure 3.b presents the Call Blocking Percentage as a function of the Call Dropping Percentage. According to this figure, the two schemes achieve the same CBP for the same CDP, with an advantage of DCAC over GC when the CDP is bellow 10%. It is worth noting that typical interesting CDP values are bellow 10%. Figure 3.c depicts the number of finished users as a function of the CDP. Here again, the DCAC scheme allow more users to finish their calls when the CDP is bellow 10%.

Figure 3.d presents the mean holding time of dropped users as a function of the CDP. The mean holding time presented in this figure is computed for only dropped users and

reflects the mean holding time for users who were not able to finish their calls. As presented in this figure the DCAC is able to protect users in the network from being dropped. Not only the scheme is able to achieve better performance in terms of CDP and resource utilization but also to increase the mean holding time before being dropped. This means that all users with shorter holding time are saved from being dropped.

The obtained results of this comparison can be explained by the fact that the DCAC scheme is able to differentiate between those data users who are most likely to be dropped and those who are not. By denying access to those data users who have responses bellow the acceptance threshold even if the resources are available locally, the DCAC scheme protects the other users from being dropped. However, the GC scheme is unable to make this differentiation and data users with high dropping probability get access to the network. Those users consume resources which may prevent other users from surviving a handoff.

The improvement of using the DCAC when only data users are present is not very big however, results show that even with one traffic type the DCAC achieves better performance. In the next subsection, we compare the two schemes when the three types of traffic are present.

4.2.2 Second set of simulations: all classes of service

In this set of simulations we simulate a system where users can have one of three types of traffic; voice, data or video. The probabilities associated with these types are $R_{voice} = 0.3$, $R_{data} = 0.4$ and $R_{video} = 0.3$ respectively.

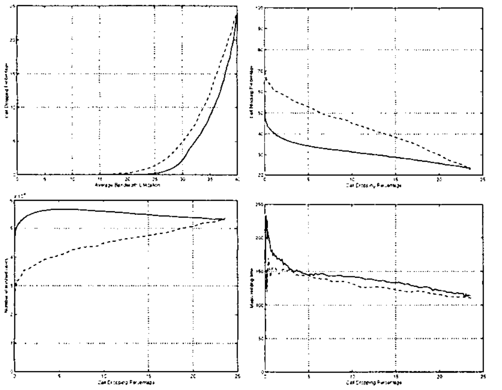

Figure 4. (4.a, 4.b, 4.c and 4.d) Comparing the two schemes (all classes of service)

Figure 4 presents the comparison results of the two schemes when the three types of traffic are present. Solid lines refer to the DCAC scheme and dotted lines refer to the GC scheme.

Figure 4.a depicts the Call Dropping Percentage experienced by all users according to the average bandwidth utilization achieved by the two schemes. As presented by figure 4.a, for the same average bandwidth utilization the DCAC scheme allows for a lower CDP. The DCAC is able to achieve the same CDP while allowing a higher resource utilization. We can notice that the improvement in this case is higher than in the case of data traffic only (in figure 3.a).

Figure 4.b presents the Call Blocking Percentage as a function of the Call Dropping Percentage. According to this figure, the DCAC scheme achieves a very low CBP for the same CDP value. This is a very important result, since it shows the importance of differentiating between users in an admission control scheme. Being able to differentiate between users allow the DCAC to accept more users without sacrificing the CDP. In the same case, to achieve the same CDP the GC scheme has no choice but to block more users. This is better shown in figure 4.c which depicts the number of finished users as a function of the CDP for the two schemes. According to this figure, it is clear that the DCAC scheme achieves better performance as a very high number of users are able to finish their calls without being dropped as opposed to the GC case.

Figure 4.d presents the mean holding time of dropped users as a function of the CDP. Here again, the DCAC scheme achieves better performance.

To better understand the behavior of the two schemes, let us have a look at the performance of the two schemes with regards to the three type of traffic separately.

Figure 5 depicts the performance of the two schemes for each type of traffic. Figure 5.a presents the number of finished voice users according to the CDP. As the GC scheme requires a higher number of reserved channels to get a lower CDP, the number of finished voice users decreases while the CDP decreases. However, for the DCAC scheme, the number of finished voice users increases while the CDP decreases. At first sight, this result may seam abnormal since the number of finished voice users is expected to decrease as the CDP decreases (i.e. as the AT increases). However, as the AT increases, less data and video users are accepted to the network which gives voice users more chance to get accepted and to finish their calls without being dropped.

Figure 5.b depicts the Call Blocking Percentage experienced by voice users according to the CDP. As shown by this figure, voice users experience a lower CBP with DCAC than with GC for the same CDP.

Figure 5.c presents the number of finished data users according to the CDP. As expected, for the GC scheme, the number of finished data users decreases as the CDP decreases. However, the DCAC scheme presents an interesting behavior. As the CDP decreases, the number of finished

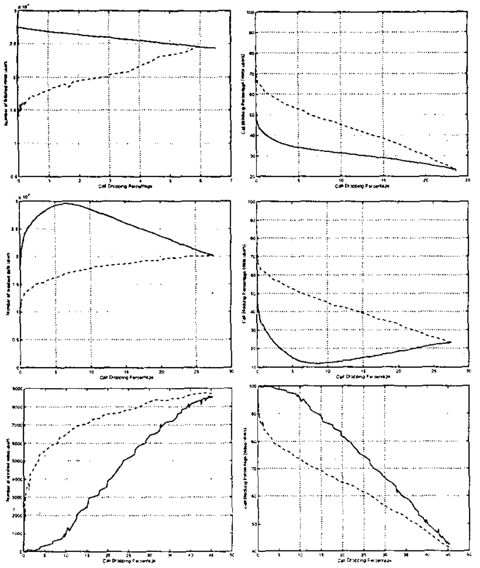

Figure 5. (5.a, 5.b, 5.c, 5.d, 5.e and 5.f) Comparing the two schemes (all classes of service)

data users increases and then decreases. The increase in the number of finished data users can be explained by the fact that as CDP decreases (i.e. AT increases), less video users are accepted to the network. This gives data users more chance to get accepted to the network and finish their calls. However, as the CDP gets smaller and smaller, the DCAC can no longer protect data users and will then block more data users from accessing the network which explains why the number of finished data users decreases when the CDP value gets bellow 7%. Figure 5.d depicts the Call Blocking Percentage experienced by data users according to the CDP. Here again, for the GC scheme, the CBP experienced by data users increases while CDP decreases. But for the DCAC scheme, the same phenomena as presented in figure 5.c happens. The CBP for data users decreases as the CDP decreases. This is because, when the AT increases, less video users are accepted to the network which gives data users more chance to get accepted and finish their calls. However, when the CDP reaches a certain value, no more data users can be accepted to the network without sacrificing voice users, so the CBP for data users start increasing while the CDP decreases.

Figure 5.e and 5.f respectively presents the number of finished video users and the CBP for video users according to the CDP. These two figures show that the DCAC scheme is able to detect the congestion and start blocking video users that will most likely be dropped due to their high bandwidth demand (10 BUs). The GC scheme, on the other hand, will still allow more video users to access the network at the expense of a higher CDP value and at the expense of other voice and data users. It is worth noting that the simulations conducted assume a cell load of 100% which means that the network is congested.

Similar comparative simulations were conducted with a cell load of 50% and similar results were obtained.

5 Conclusion

In this paper, we have presented a performance evaluation of a distributed call admission control scheme. The call admission process involves the cell that receives the call admission request and a cluster of neighboring cells. This distributed process allows the original cell to make a more clear-sighted admission decision for the new user.

Simulations results have shown a significant improvement of the distributed CAC over the Guard Channel scheme. By implementing the distributed call admission control scheme, the system is able to lower the call dropping probability while offering a high resource utilization by selecting only those users that are more likely to complete their call without being dropped.

References

[1] S. Choi and K. G. Shin. Predictive and adaptive bandwidth reservation for hand-offs in qos-sensitive cellular networks. *in Proc. ACM SIGCOMM'98*, pages 155–166, September 1998.

[2] J. Evans and D. Everitt. Effective bandwidth based admission control for multiservice cdma cellular networks. *IEEE Trans. Vehicular Tech.*, 48(1):36–46, January 1999.

[3] R. Guerin. Queueing blocking system with two arrival streams and guard channels. *IEEE Trans. on Comm.*, 36:153–163, 1988.

[4] D. Hong and S. Rappaport. Traffic modeling and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures. *IEEE Trans. Vehicular Tech.*, 35:77–92, 1986.

[5] Y. Iraqi and R. Boutaba. A novel distributed call admission control for wireless mobile multimedia networks. *in Proc. ACM WoWMoM'2000*, pages 21–27, 2000.

[6] D. A. Levine, I. F. Akyildz, and M. Naghshineh. The shadow cluster concept for resource allocation and call admission in atm-based wireless networks. *in Proc. ACM MOBICOM'95*, pages 142–150, 1995.

[7] D. A. Levine, I. F. Akyildz, and M. Naghshineh. A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept. *IEEE/ACM Trans. on Net.*, 5(1), February 1997.

[8] S. Lu and V. Bharghavan. Adaptive resource management algorithms for indoor mobile computing environments. *in Proc. ACM SIGCOMM'96*, pages 231–242, August 1996.