A Dynamic Call Admission Scheme for VBR traffic in ATM networks

Srinivasan Ramaswamy

Newbridge Networks Corporation
600 March Road, P.O. Box 13600
Kanata, Ontario, CANADA K2K 2E6
Tel: (613) 591-3600
srini@newbridge.com

Pawel Gburzynski

University of Alberta
Department of Computing Science
Edmonton, Alberta, CANADA T6G 2H1
Tel: (403) 492-2347
pawel@cs.ualberta.ca

Abstract: The role of call admission control (CAC) in high-speed networks is to maintain the network utilization at a high level, while ensuring that the quality of service (QoS) requirements of the individual calls are met. We propose a generic dynamic call admission scheme for VBR and ABR traffic whose aim is to reduce the blocking rate for VBR calls at the expense of a higher blocking rate for ABR calls. Our scheme is generic because it builds up on a pre-existing static scheme, e.g., one based on a simple notion of *effective bandwidth*.

Keywords: call admission, bandwidth characterization, ATM networks

1.Introduction

We use the term static CAC to describe schemes that always allocate the same bandwidth to a specific group of multiplexed calls, independent of the other traffic sharing the link. Such schemes are usually based on the concept of *effective bandwidth* [1, 5, 10]. Dynamic CAC, on the other hand, denotes a scheme in which the bandwidth allocation to a group of calls sharing a queue is influenced by the traffic in other queues destined for the same outgoing link.

We propose a generic dynamic call admission scheme (D-CAC) for VBR and ABR traffic, whose aim is to reduce the blocking rate for VBR calls (α_{VBR}) at the expense of a higher blocking rate for ABR calls. The rationale behind this approach is that the VBR traffic is considered more critical than ABR traffic, and, generally, it will bring more revenue to the network provider.

Our system consists of an output port at an ATM switch. The two incoming traffic streams, i.e., VBR and ABR traffic, are multiplexed into two queues. A weighted fair queuing [3] or a similar mechanism is used to guarantee each queue the bandwidth allocated to it by CAC. Bandwidth unused by VBR is utilized by ABR traffic.

Calls being multiplexed in the same queue are assumed to have the same QoS requirements. ABR calls have a minimum bandwidth requirement while the bandwidth requirement of a VBR call—the effective bandwidth—can lie between its peak and average cell rate. Switch buffers are divided among the output ports. The buffers allocated to an output port are statically apportioned among the two queues.

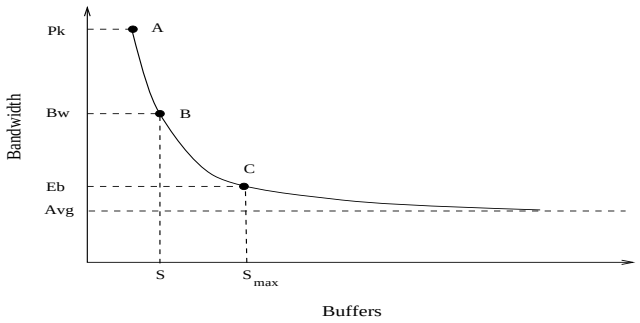

Figure 1: Resource tradeoff plot for a VBR source

2.Bandwidth-buffer tradeoff

Two resources, link bandwidth and buffers, are shared by the two traffic classes. Since VBR traffic is delay-sensitive while ABR traffic is sensitive to cell loss, it should be possible to allocate a larger share of the bandwidth than strictly necessary to the former—giving the queue for ABR traffic a large buffer size—and still meet the QoS needs for both classes. Therefore, there is more than one possible bandwidth-buffer allocation that will meet the desired QoS for the two traffic classes. A hypothetical tradeoff plot is shown in figure 1. To obtain it, the VBR delay and loss rate for a given buffer size and service rate can be calculated using, for example, the analytical model in [1].[1] The tradeoff plot can also be calculated using the effective bandwidth schemes described in [5, 10].

When the size of the queue is small, a large bandwidth allocation—close to the peak rate (point A)—is needed to satisfy the delay and cell loss require-

[1]The model predicts the queue overflow probability alone. However, we have extended it to predict average delay as well [8].

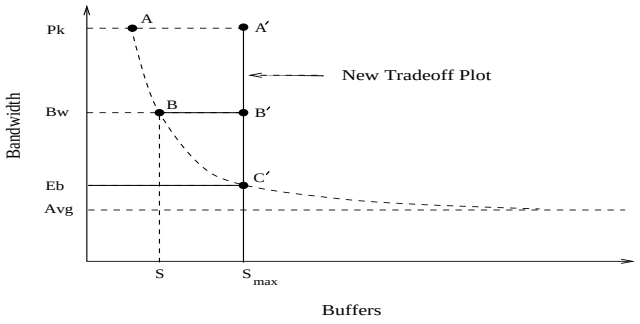

Figure 2: Tradeoff plot for a VBR call (fixed buffer size)

ments. However, as more buffers become available, the bandwidth needed to satisfy the QoS requirements reduces asymptotically to the average rate. In practice, VBR traffic has a maximum delay/jitter bound that will restrict the buffer size to a maximum value, S_{max}. The bandwidth corresponding to this buffer size on the tradeoff plot (point C), is called the effective bandwidth, EB. A static scheme would use this value as the bandwidth requirement of an incoming VBR call.

3.Dynamic CAC

We would prefer to retain the advantages of dynamically relocating the operating point without the need to trade cell buffers back and forth between the two queues. If we fix the buffer size of the VBR queue at S_{max}, then allocating peak bandwidth to the VBR call would lower its average delay *as well as* its cell loss rate. This is an unnecessary over-allocation, as VBR traffic is not as sensitive to cell loss as it is to delay and delay jitter.

To resolve this difficulty, one can employ a feedback scheduler for the VBR queue [6]. The service rate of the VBR queue is modulated around a nominal allocation made by CAC in order to reduce the effect of traffic fluctuations and traffic parameter variations on the QoS. The WFQ mechanism guarantees the VBR traffic its share of the bandwidth; as before, any remaining bandwidth can be utilized by the ABR queue. The usefulness of this scheduling mechanism stems from the fact that irrespective of the service rate allocated by CAC, the cell loss rate is maintained at the requisite level. The service rate allocation only affects the average delay, as shown in [8].

This means that CAC can make any service rate allocation and know that the cell loss rate will be close to the requirement. Allocating close to peak rate will lower the average delay while a service rate close to effective bandwidth will result in an average delay close to the QoS specification. Thus, the combination of D-CAC and feedback control scheduling allows us to lower the call blocking rate and average delay for VBR traffic, without the buffer manipulation entailed by the use of the tradeoff plot.

Figure 2 shows the new tradeoff plot along which the operating point is varied. The operating points A' and B' lie above the old tradeoff plot. This region denotes an over-allocation of resources for a given QoS. However, because of the feedback control scheduler, the cell loss rate is maintained at the required level. Hence, only the other component of the QoS, the average cell delay, is lowered as a result of the choice of operating point.

We select a very simple static CAC scheme as the basis for the D-CAC. The effective bandwidth EB is assumed to be constant, irrespective of the number of VBR calls being multiplexed. As we only intend to compare the performance of the D-CAC against that of its prerequisite S-CAC, the accuracy of the S-CAC scheme is not important.

We use the following notation:

- L — the link rate (cells/sec) at the output port, shared between the VBR and ABR queues
- Pcr — the peak rate of a VBR call in cells/sec
- Scr — the average rate of a VBR call in cells/sec
- S_{max} — the maximum size of the VBR queue, obtained by a consideration of the maximum per-node delay and delay jitter bounds
- EB — the effective bandwidth of a VBR call in cells/sec, required to satisfy the QoS when the queue size is S_{max}
- Mcr — the minimum bandwidth requirement of an ABR call in cells/sec
- Vbr_Bw — the service rate allocated to the VBR queue; for S-CAC, it is given by the product of the number of VBR calls in progress and EB
- Abr_Bw — the service rate allocated to the ABR queue, given by the product of the number of ABR calls in progress and Mcr.
- $Avail$ — the portion of link bandwidth currently unused by either queue
- N — the number of VBR calls currently admitted to the queue
- α_{VBR} — the call blocking rate for VBR traffic
- α_{ABR} — the call blocking rate for ABR traffic

Figure 3 shows a flowchart of the D-CAC algorithm. When a call is accepted, the service rate allocated to the VBR queue can be normalized w.r.t. the number of calls in progress to obtain the dynamic effective bandwidth. This quantity lies between the static effective bandwidth EB and Pcr. In order to compare the static and dynamic CAC, we define the gain G due to D-CAC as

$$G = \frac{Vbr\_Bw}{N * EB},$$

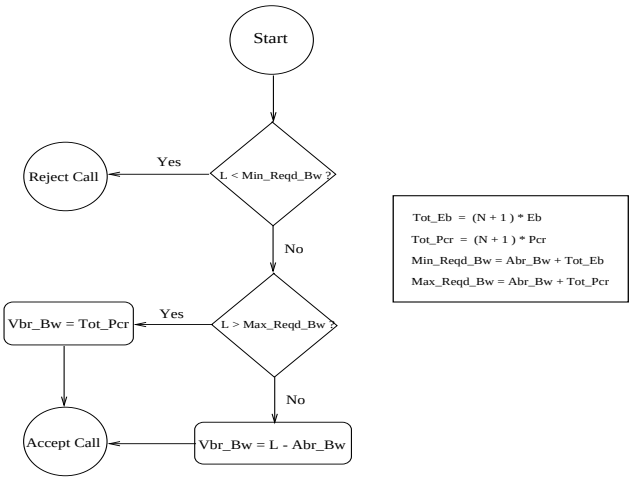

Figure 3: D-CAC – incoming VBR call

where Vbr_Bw is the service rate for the VBR queue and N is the number of VBR calls in progress.

As a result of this definition, G lies between 1 and Pcr/EB. The static CAC always has $G = 1$. If the average value of G—after a large number of VBR calls have arrived and departed from the queue—is close to Pcr/EB, it indicates that the D-CAC on average allocated each VBR call a bandwidth close to Pcr.

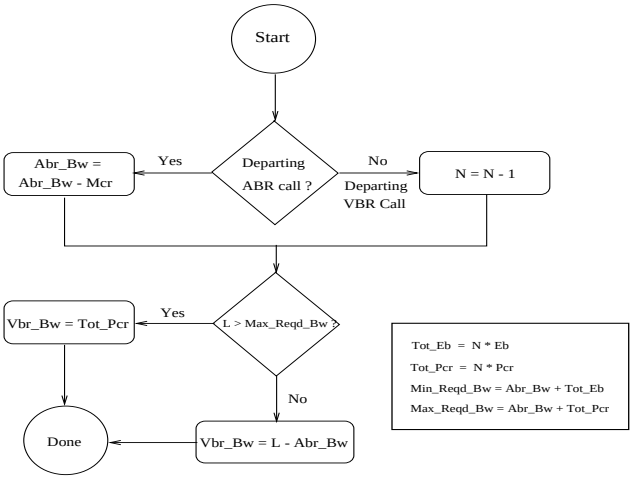

Figure 4: D-CAC – departing VBR/ABR call

When a VBR (or ABR) call departs, S-CAC merely decrements the amount of bandwidth allocated to the queue, Vbr_Bw (or Abr_Bw). However, D-CAC recalculates the amount of bandwidth allocated to the VBR queue at *every* call departure according to the algorithm shown in figure 4. Thus, while S-CAC does not allow VBR traffic to benefit by the resources freed by a departing call, D-CAC immediately uses the resources freed up to the advantage of VBR traffic.

4. Sample results

The performance of the D-CAC was evaluated via call-level simulations. Cell traffic was not generated. The following factors were considered:

ρ the ratio of the total offered load to the link capacity

Pcr/L the ratio of the peak cell rate of VBR traffic to the link rate

EB/Pcr the ratio of the static effective bandwidth of a VBR call to its peak cell rate

ρ_{abr}/ρ_{vbr} the ratio of the offered ABR load to the VBR load

The Bellcore document [4] specifies the behavior of test sources used to evaluate the performance of broadband switching systems. VBR test sources are characterized as two-state (On/Off) Markov processes. Three types of VBR test sources conforming to this characterization are shown in table 1. The On and Off periods

Source	M_a	M_s	P	Scr/Pcr	Pcr/L
VBR I	240	720	6	0.25	0.167
VBR II	500	2500	25	0.16	0.04
VBR III	210	2500	1	0.078	1

Table 1: Bellcore VBR test source parameters

are assumed to be geometrically distributed with mean durations of M_a and M_s cells respectively. During an On period, a cell is generated every P cell slots.

The levels of factors chosen for our experiments are listed in table 2. From table 1, it can be seen that

Factor	**Levels**
ρ	0.6, 0.7, 0.8, 0.9
Pcr/L	0.04, 0.4
EB/Pcr	0.2, 0.6
ρ_{abr}/ρ_{vbr}	0.1, 1, 3, 7, 10

Table 2: Levels of factors used in experiments

Scr/Pcr of the VBR source types lies in [0.08, 0.25]. Therefore, in choosing the levels for EB/Pcr, we pick a value close to Scr/Pcr—corresponding to a bandwidth allocation of average rate, and another value (closer to 1)—corresponding to near peak rate allocation.

Figure 5 shows the effect of the ratio $EB/L = (Pcr/L) * (EB/Pcr)$, for two values of EB/Pcr. Pcr itself has no effect on the call blocking rate (CBR) when S-CAC is used. It can be seen that the call blocking rate

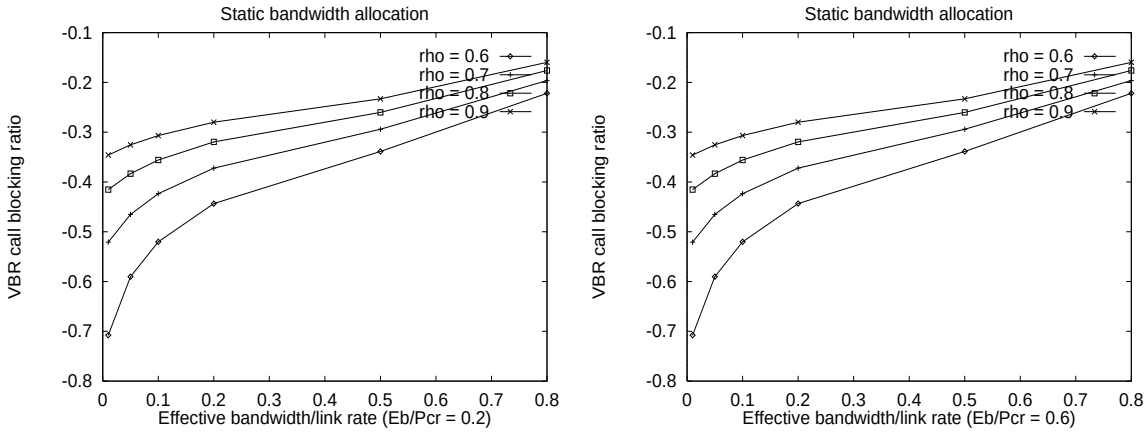

Figure 5: S-CAC: effect of EB/L [note that the second figure is the same as the first one]

increases as the effective bandwidth approaches the link rate. This is because, for a given load and call arrival rate, *calls with smaller bandwidth requirements have a higher chance of being accepted.* Also, for calls with small EB, the load factor ρ has a considerable effect on the call blocking ratio. As expected, the EB/Pk ratio does not have any effect on the call blocking ratio.

Figure 6 plots CBR as a function of EB/Pcr for two values of Pcr/L. As expected, the peak rate has no effect when static CAC is used.

Next, we investigate the effect of ρ_{abr}/ρ_{vbr} on call blocking. In the first set of experiments, Pcr/L was fixed at 0.04 and $EB/Pcr = 0.2$ and 0.6. The experiments were repeated with $Pcr/L = 0.4$. Since Pcr has no effect on the performance of S-CAC, this is equivalent to conducting the experiments with four levels of EB/L. Figure 7 shows the results for $Pcr/L = 0.04$. When the ratio of load offered by ABR traffic to that offered by VBR traffic is increased—keeping the call arrival ratio fixed—the bandwidth requirement Mcr of an ABR call increases. This increases the call blocking rate for ABR—just as the call blocking rate for VBR increases with EB in figure 5 (left). Therefore, for a given total load, the call blocking rate for VBR decreases as ρ_{abr}/ρ_{vbr} increases. Figure 8 shows the results for $Pcr/L = 0.4$. The conclusions drawn for $Pcr/L = 0.04$ still apply, although the blocking rates are higher due to the higher bandwidth requirements. It can be seen that as EB approaches the link rate, the call blocking rate flattens off (right).

Next, we use D-CAC and compare the effect of the parameters on the call blocking rate for VBR traffic. Figure 9 shows the effect of EB/L. The shape of the plots are the same as for S-CAC, but the D-CAC plots have a sharper knee. This suggests that at high EB (more than 1/5 of the link rate), there may not be much advantage in using the dynamic scheme.

It is clear by comparison with figure 5 that the call blocking rate for low EB/L is at least an order of magnitude lower with D-CAC. The blocking rate increases with the bandwidth requirements. For calls with EB close to the link rate, the D-CAC call blocking rate approaches that for S-CAC.

Figure 10 shows the effect of EB/Pcr. While the performance of S-CAC was unaffected by the peak rate (for a given effective bandwidth), EB/Pcr can be seen to have a large impact on the performance of D-CAC, especially at medium loads. The figure shows that the blocking rate increases with EB/Pcr. This is because the amount of bandwidth above EB that can be captured by a VBR call decreases with increasing EB/Pcr. Therefore, we can conclude that D-CAC works well when the effective bandwidth requirement is small compared to the link rate and the peak-to-average ratio is high, i.e., the traffic is bursty.

Finally, we investigate the effect of ρ_{abr}/ρ_{vbr} on call blocking. As in the case of S-CAC, the first set of experiments were conducted with $Pcr/L = 0.04$ and $EB/Pcr = 0.2$ and 0.6. The experiments were then repeated with $Pcr/L = 0.4$. This was not the same as conducting the experiments at four levels of EB/L, since the EB/Pcr ratio impacts the performance of D-CAC (but not S-CAC).

Figures 11 and 12 show the results for $Pcr/L = 0.04$ and $Pcr/L = 0.4$, respectively. At low EB/Pcr, the call blocking rate is unaffected by the ratio of ABR load to VBR load (see the left parts of figures 11 and 12). At high EB/Pcr, the call blocking rate increases with ABR load (the right parts of the two figures) as in the case of S-CAC (the reason for this increase was explained before).

At low EB/Pcr, bandwidth reservation for a VBR call close to the peak rate results in a large degree of over-allocation (greed). Even at low ρ_{abr}/ρ_{vbr} (or low

bandwidth requirements for ABR calls), the high greediness causes the ABR calls to be rejected. The only reason in this case for VBR call blocking is the competition among VBR calls. Therefore, when the ρ_{abr}/ρ_{vbr} ratio increases—resulting in higher bandwidth requirements for ABR calls and correspondingly higher chances of blocking—there is no change in the VBR blocking ratio.

At high EB/Pcr, a reservation close to peak rate does not yield as much benefit—the greediness is small. Hence, when the bandwidth requirements for ABR calls are low, VBR calls are rejected as a result of competition with ABR calls. When the bandwidth requirements for ABR calls increase (due to ABR load increase), the blocking rate for ABR increases leading to an improvement in the blocking rate for VBR.

5.The modified D-CAC algorithm

Table 3 compares the VBR call blocking rate (α_{VBR}) for D-CAC and S-CAC. The calls are assumed to be bursty ($EB/Pcr = 0.2$) and the degree of statistical multiplexing is high ($Pcr/L = 0.01$). The blocking rate for ABR (α_{ABR}) is also shown. These results indicate

Load ρ	**S-CAC**		**D-CAC**	
	α_{VBR}	α_{ABR}	α_{VBR}	α_{ABR}
0.6	0.2	0.2	3.9997e-07	0.99
0.7	0.3	0.3	1.5e-04	0.99
0.8	0.38	0.38	4e-03	0.99
0.9	0.45	0.45	2.7e-02	0.99

Table 3: D-CAC vs. S-CAC

that D-CAC is very effective in reducing the call blocking rate for VBR traffic—unfortunately, ABR traffic is very severely affected. As a remedy, we allow the user to specify the maximum call blocking rate allowable for ABR traffic (α_{ABR}^{max}). D-CAC is expected to improve the call admission ratio for VBR traffic as long as the call blocking rate for ABR traffic does not exceed the specified value.

A simple modification to the D-CAC algorithm (the resulting scheme is termed MD-CAC) proved to yield satisfactory results. When the CBR for ABR traffic, α_{ABR}, exceeds $0.8 * \alpha_{ABR}^{max}$ (upper threshold), the maximum dynamic effective bandwidth is set to be equal to EB—equivalently, the dynamic gain $G = 1$—i.e, the queue is allocated a service rate identical to the S-CAC allocation. This conservative allocation allows new ABR calls to be accepted, lowering the call blocking rate for ABR traffic.

When α_{ABR} falls below $0.5 * \alpha_{ABR}^{max}$ (lower threshold), the maximum dynamic effective bandwidth for the VBR queue is set equal to the sum of the peak rates of the VBR calls (dynamic gain $G = Pcr/EB$)—which means that the MD-CAC is free to allocate any bandwidth between $\sum_{i=1}^{N} EB$ and $\sum_{i=1}^{N} Pcr$ to the VBR queue. This hoarding of bandwidth by the VBR calls may cause new ABR calls to be turned away, raising α_{ABR}.

Figures 13 and 14 show the behavior of MD-CAC as the desired blocking rate for ABR, α_{ABR}^{max}, is varied. The maximum gain is achieved in figure 13 (left), showing that MD-CAC is most effective for bursty calls with bandwidth requirements much smaller than than link rate. When EB is close to Pcr, not much benefit can be expected from MD-CAC (which is natural). Figures 15 and 16 plot the achieved ABR call blocking ratio, α_{ABR}, against the desired value, α_{ABR}^{max}. In all cases, the actual blocking rate is seen to be bounded by the upper and lower limits specified.

Compared to S-CAC, D-CAC and MD-CAC improve the call blocking rate for VBR traffic at the expense of ABR traffic. Hence, the blocking rate for ABR cannot be lower than that achieved by S-CAC. When the desired blocking rate is set below the S-CAC value, it cannot be achieved. This accounts for the initial flat portion of the curves lying outside the desired limits. Once the desired blocking rate exceeds the minimum value obtained by use of S-CAC, the performance of D-CAC begins to improve, with a corresponding increase in ABR blocking rate.

As mentioned earlier, the dynamic gain G allows us to compare the performance of static and dynamic CAC. Figures 17 and 18 plot G as a function of the desired ABR call blocking rate. Since the dynamic gain normalizes the dynamic effective bandwidth w.r.t. EB, the dynamic gain comparison can only be made for plots with the same EB. Thus, in figure 17, we see that higher gain is achieved for smaller EB/Pcr (greater burstiness).

When $EB/Pcr = 0.6$ ($Pcr/EB \approx 1.5$), we see that G reaches the maximum value of 1.5 (for $\rho = 0.6$), indicating that VBR calls are being allocated peak rate on average. In spite of this, the call blocking rate is higher (figure 13) than for $EB/Pcr = 0.2$ (in the latter case, the gain is 1.7 against a maximum possible gain of 5 with the average allocation of one-third of the peak rate). This shows that it is the gain G, rather than the average allocated bandwidth, that is an index of the improvement in performance as a result of MD-CAC.

From figure 17 (left), we see that when $\rho = 0.9$, the maximum gain is 1.1 (at $\alpha_{ABR}^{max} = 1$) which means that the average bandwidth allocated by MD-CAC is $1.1 * EB$. From the corresponding plot in figure 13, we see that the call blocking rate is lower by more than an order of magnitude than in the static case. This shows that, for bursty calls, a small degree of over-allocation in bandwidth can result in a marked improvement in

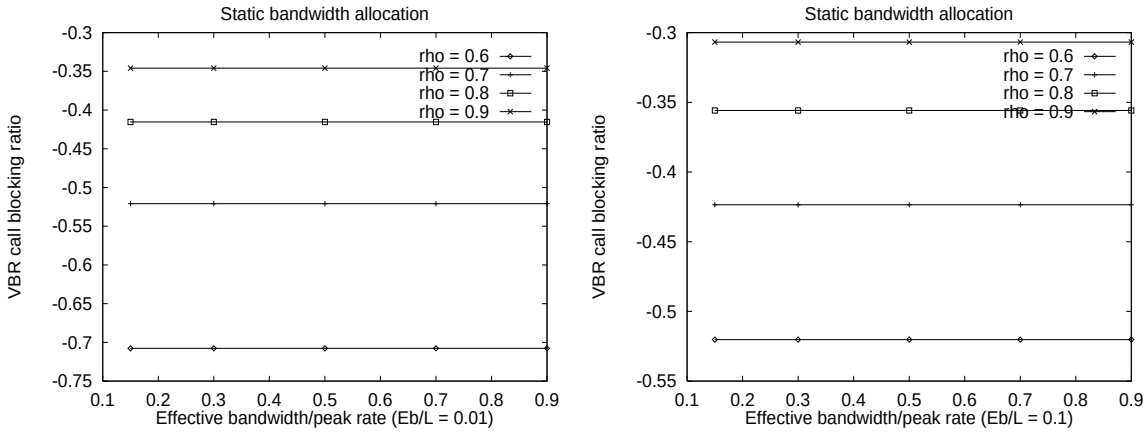

Figure 6: S-CAC: effect of EB/Pcr

the call blocking rate.

Comparing figures 17 and 13, we can see that when the dynamic gain is high, the call blocking rate is correspondingly reduced. A similar conclusion can be drawn from figures 18 and 14, which goes to prove that the dynamic gain is a good indicator of the improvement resulting from the use of MD-CAC—as noted earlier.

6. Conclusions

We propose a dynamic CAC scheme aimed at improving the call blocking rate for VBR traffic, at the expense of a higher blocking rate for ABR calls. We show that the dynamic scheme is most beneficial for bursty calls whose individual bandwidth requirements are small compared to the link rate. VBR voice traffic is an important member of this class. For a high-speed link, this property will hold for all bursty traffic patterns, including video, teleconferencing, etc. The call blocking rate was shown to improve by several orders of magnitude over static allocation in such circumstances. The impact of the dynamic scheme on ABR blocking can be restricted by specifying the maximum acceptable ABR blocking rate.

The proposed scheme is simple and easy to implement. It can be used in conjunction with any static bandwidth allocation scheme and not just the fixed bandwidth allocation scheme described in this study. Since the dynamic CAC is invoked at call departure as well as call arrival, it allows existing VBR calls to benefit from the resources freed by the departing call—unlike static allocation schemes.

References

[1] D. Anick, D. Mitra, and M. M. Sondhi. Stochastic theory of data–handling system with multiple sources. *The Bell Technical Journal*, 61(8):1871, 1982.

[2] W. W Armstrong and M. Thomas. *Handbook of Neural Computation – Adaptive Logic Networks(Section C1.8).* Oxford University Press, 1996.

[3] A. Demers, S. Keshav, and S. Shenker. Analysis and simulation of a fair queueing algorithm. *Journal of Internetworking Research and Experience*, pages 3–26, October 1990.

[4] GR-1110-CORE. Broadband switching system (BSS) generic requirement). Technical Report Issue 1, Bellcore, September 1994.

[5] R. Guerin, H. Ahmadi, and M. Naghsineh. Equivalent capacity and its application to bandwidth allocation in high-speed networks. *IEEE Journal on Selected Areas in Communications*, 9(7):968, 1991.

[6] Srinivasan Ramaswamy and Pawel Gburzynski. An adaptive feedback control based cell scheduler for ATM networks. Submitted to Journal of High Speed Networks, 1997.

[7] Srinivasan Ramaswamy, Theodore Ono-Tesfaye, William Armstrong, and Pawel Gburzynski. A simulation approach to equivalent bandwidth characterization in ATM networks. *IASTED '96 International Conference on Modelling, Simulation and Optimization*, May 1996.

[8] Srinivasan Ramaswamy, Theodore Ono-Tesfaye, William Armstrong, and Pawel Gburzynski. Equivalent bandwidth characterization for real-time CAC in ATM networks. Submitted to Journal of High Speed Networks, 1997.

[9] Srinivasan Ramaswamy, Theodore Ono-Tesfaye, and Pawel Gburzynski. A regression approach to equivalent bandwidth characterization in ATM networks. *MASCOTS '97*, January 1997.

[10] K. M. Rege. Equivalent bandwidth and related admission criteria for ATM systems—a performance study. *International Journal of Communication Systems*, 7:181, 1994.

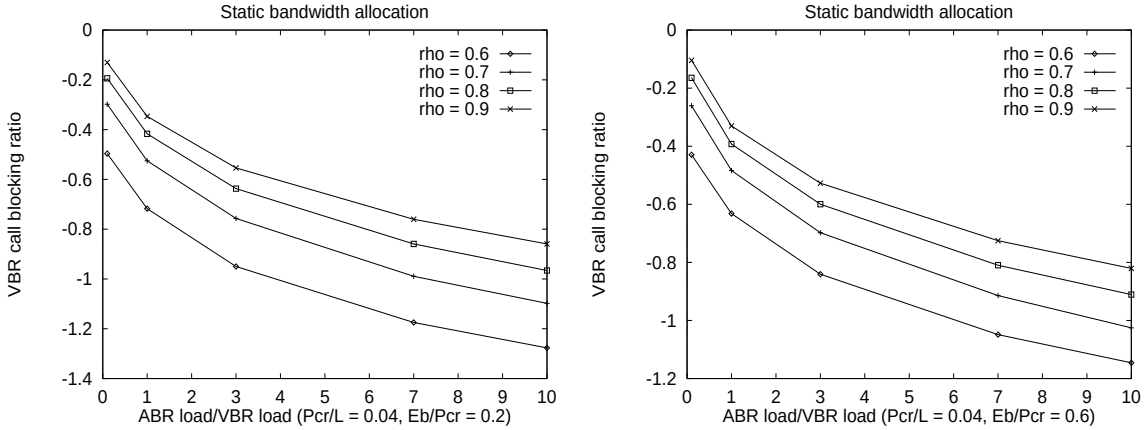

Figure 7: S-CAC: effect of ρ_{abr}/ρ_{vbr} $(Pcr/L = 0.04)$

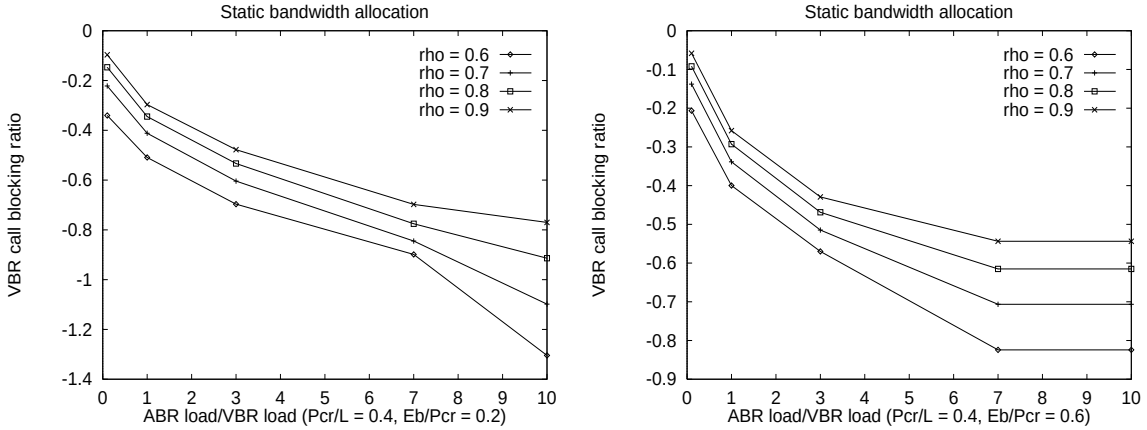

Figure 8: S-CAC: effect of ρ_{abr}/ρ_{vbr} $(Pcr/L = 0.4)$

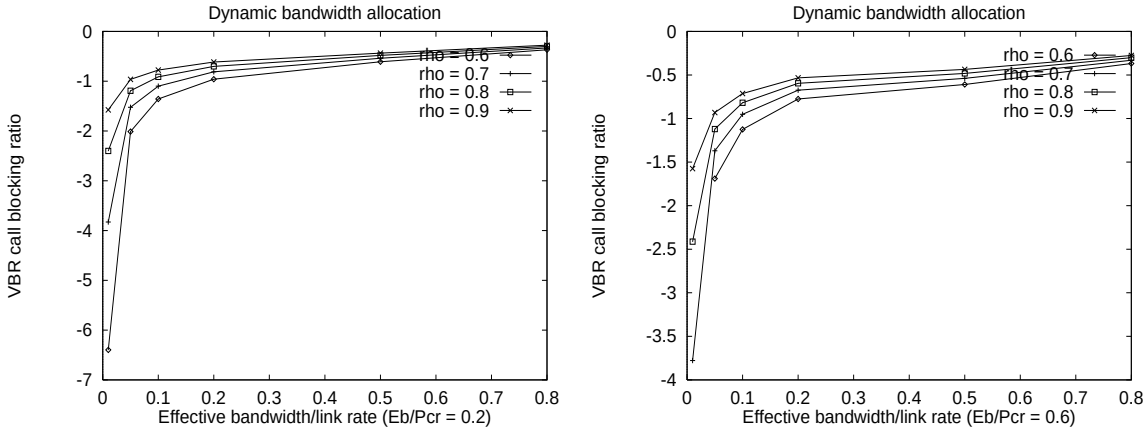

Figure 9: D-CAC: effect of EB/L

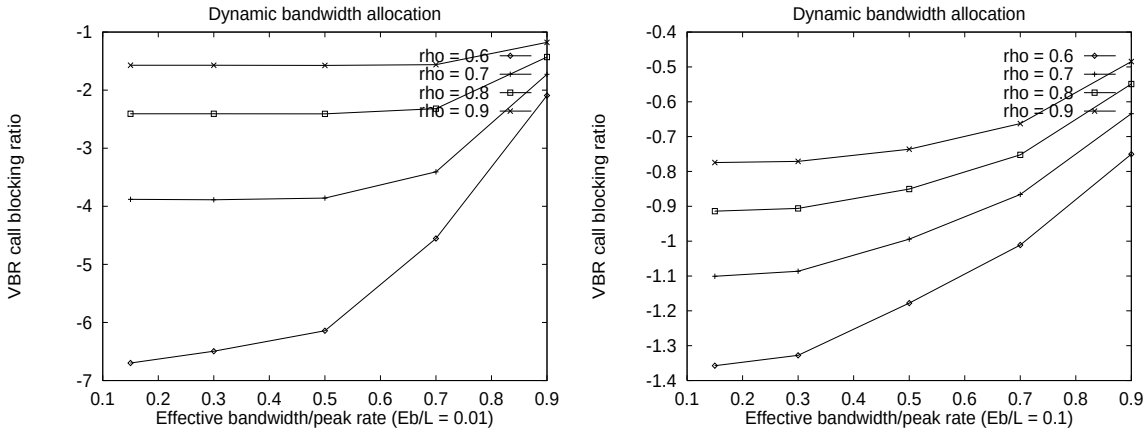

Figure 10: D-CAC: effect of EB/Pcr

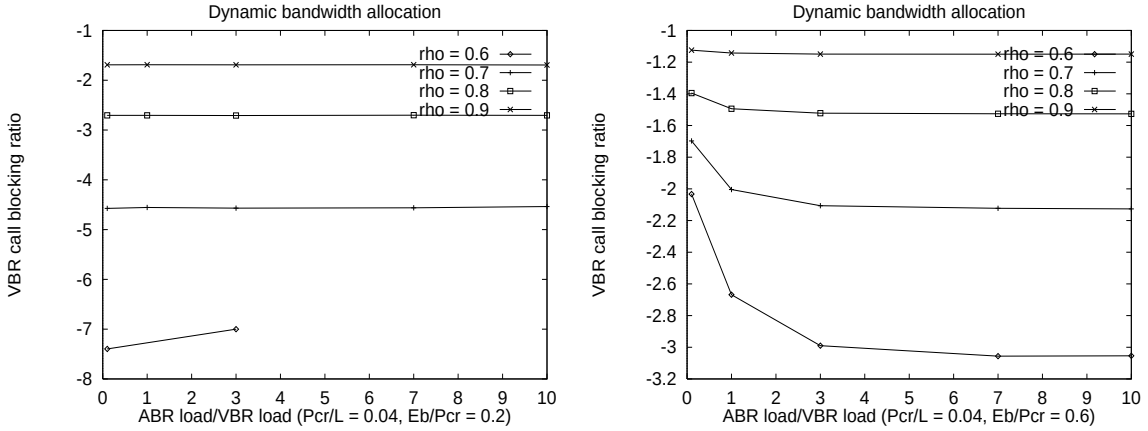

Figure 11: D-CAC: effect of ρ_{abr}/ρ_{vbr} $(Pcr/L = 0.04)$

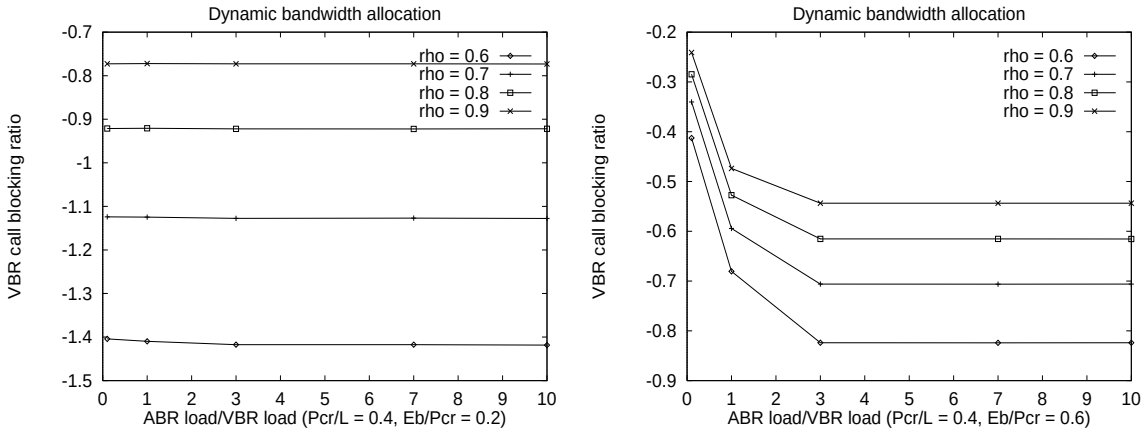

Figure 12: D-CAC: effect of ρ_{abr}/ρ_{vbr} $(Pcr/L = 0.4$

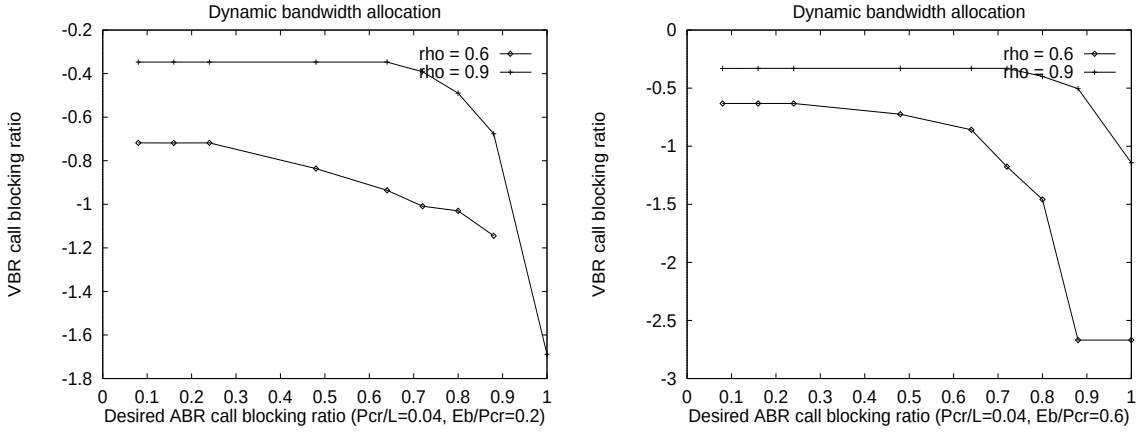

Figure 13: MD-CAC: effect of α_{ABR}^{max}

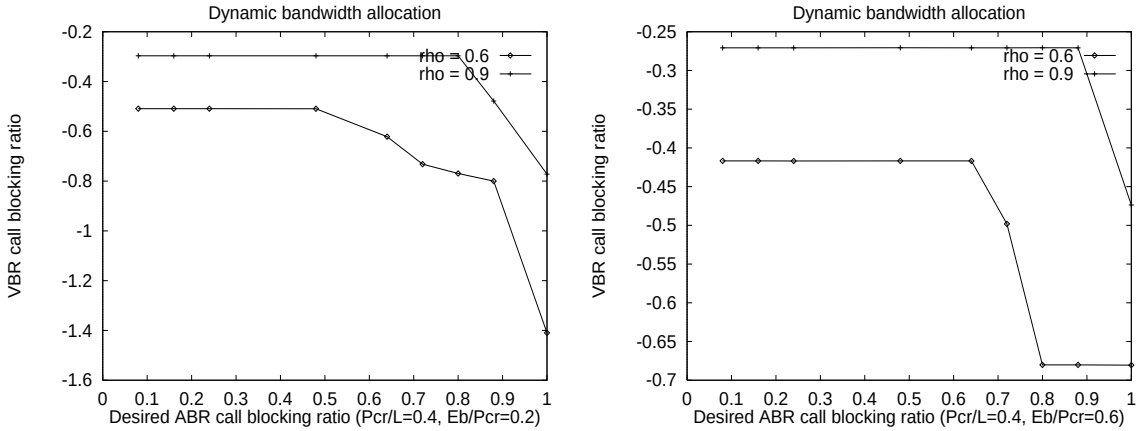

Figure 14: MD-CAC: effect of α_{ABR}^{max}

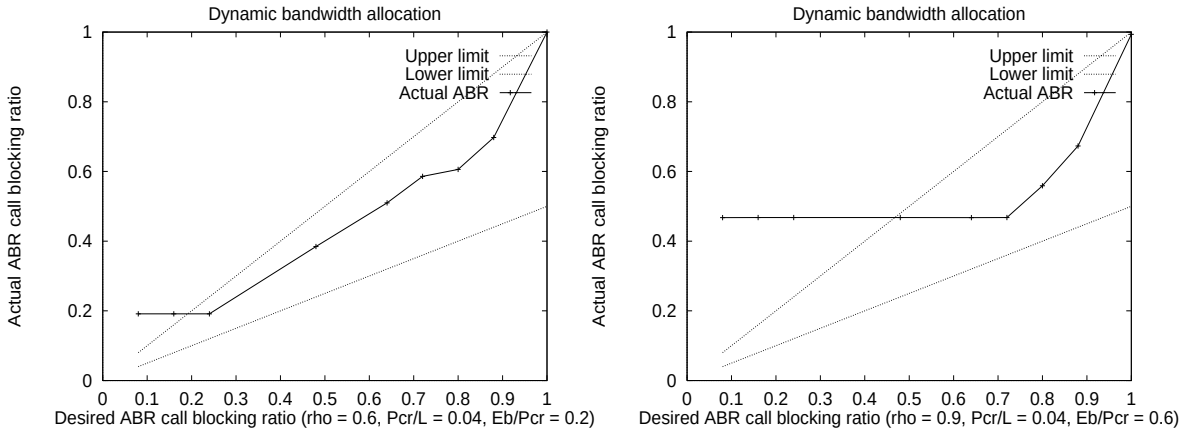

Figure 15: MD-CAC: achieved ABR blocking rate

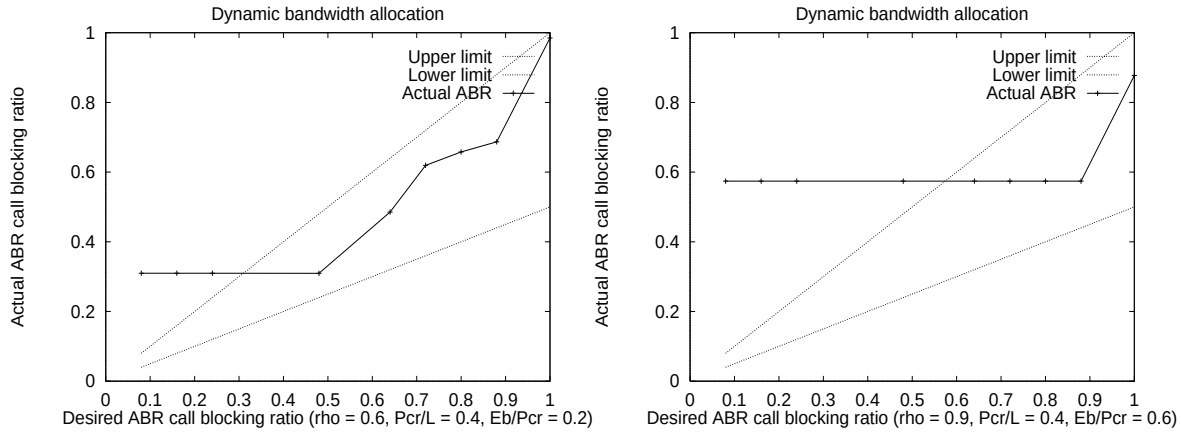

Figure 16: MD-CAC: achieved ABR blocking rate

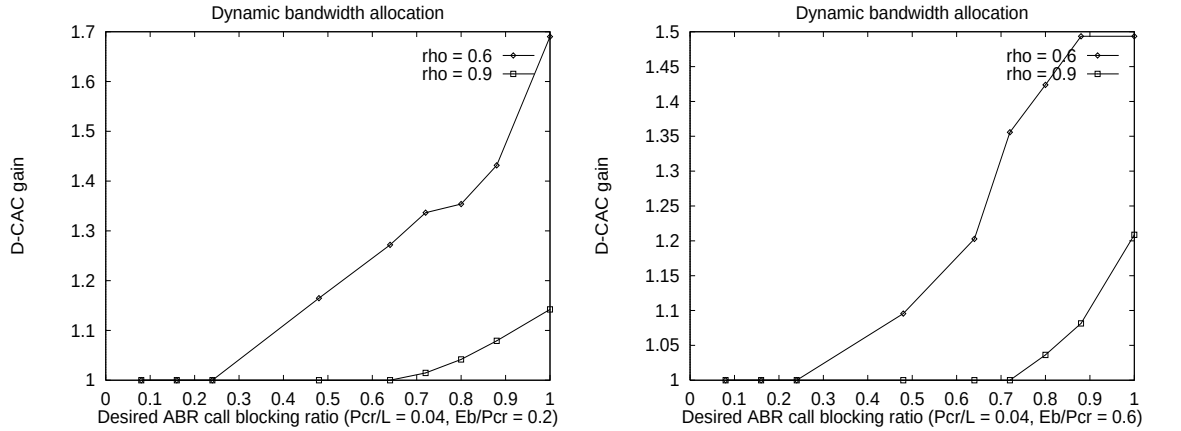

Figure 17: MD-CAC: dynamic gain

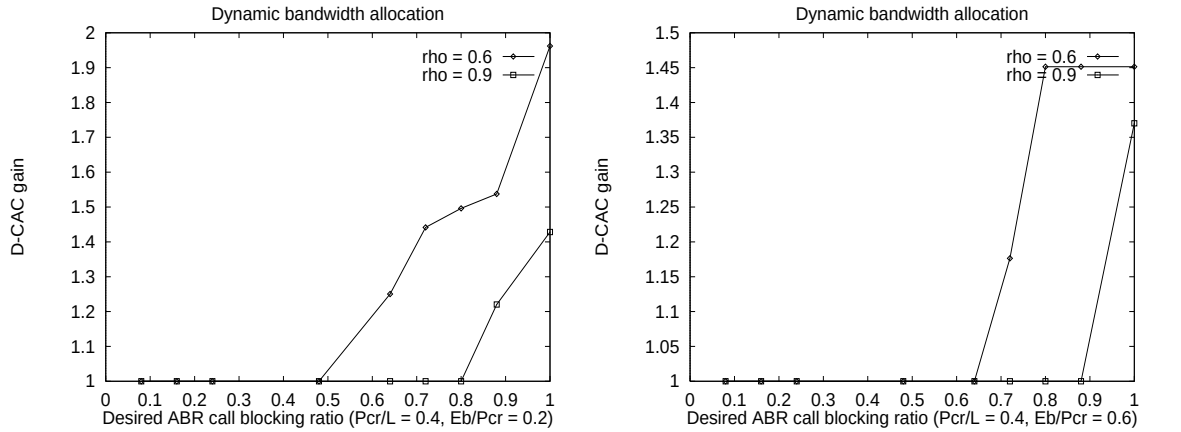

Figure 18: MD-CAC: dynamic gain