Performance Evaluation of Two CAC Algorithms in ATM Networks

Niudomar Siqueira de Araújo Chaves
Shusaburo Motoyama

Departamento de Telemática - Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas - Unicamp
Caixa Postal 6101, CEP: 13083-970, Campinas SP - Brazil
Phone: +55 19 7883765 Fax: +55 19 289-1395
{niudomar, motoyama}@dt.fee.unicamp.br

Abstract - **This paper presents a performance study of two CAC (Connection Admission Control) algorithms. Both algorithms are based on effective bandwidth concept. The results were obtained through simulation. The analysis showed that the required QoS is achieved by the two algorithms, however, they overestimate the necessary bandwidth resulting in lower network resource utilization.**

Keywords - ATM, CAC, Performance Evaluation, Simulation.

1. Introduction

The Asynchronous Transfer Mode (ATM) technology was chosen by ITU for the implementation of B-ISDN. An ATM network must be able to transport multimedia traffic. It should transport, simultaneously, several traffic types in an efficient manner. Furthermore, it should have standard interfaces clearly defined from public exchanges through local networks operating at rates from dozens through hundreds of Mbps.

The ATM technology is based on packet switching. In order to make switching and storing simpler, the packet has a fixed length and is called "cell". A cell has a 5 bytes header with switching information, followed by a payload of 48 bytes.

To simplify cell routing virtual circuit switching is used. Such connection oriented network implies a negotiation phase between user and network which results in a traffic contract. The traffic contract is described in terms of parameters like peak rate, mean rate, maximum burst size and cell delay variance.

It is common sense that multimedia traffic is essentially bursty. ATM switches use statistical multiplexing in order to use silent periods from one connection to transport cells of another active connection. The objective is an efficient link bandwidth utilization. The commitment between quality of service (QoS) and the efficient link utilization led to the development of traffic and congestion control mechanisms.

The Connection Admission Control (CAC) is a function that evaluates if a connection can be accepted or should be rejected. During the call establishment phase the user negotiates the traffic parameters with the network, which should evaluate if it can accept the requested connection and at the same time maintain the QoS of the previously established calls. This function is performed in all ATM switches along the route and the connection is accepted only if all of them can satisfy the traffic contract. The cell loss ratio (CLR) is an important QoS parameter, although the cell transfer delay and the cell delay variance are also important parameters in services like telephony and videoconference.

Several methods have been proposed for the implementation of CAC in ATM networks [1], [2], [4], [5], [6], [8], [9], [10] e [11]. Most of them are based on effective bandwidth concept [1], [2], which means the necessary bandwidth to serve only one connection separately.

Other methods use dynamic CAC schemes [4], [6], [8], [9]. The decision about the acceptance is taken based on real time traffic measurement. All methods discussed above are based on probabilistic models. Le Boudec and Nagarajan [11] proposed a deterministic method which considers the worst case cell transfer delay.

The objective of this work is to evaluate through discrete event simulation the algorithms based on effective bandwidth proposed by Guérin, Ahmadi and Naghshineh [1] (referenced as GAN within this paper), and Kesidis, Walrand and Chang [2] (referenced as KWC).

2. Effective Bandwidth Algorithms

The scheme proposed by Guérin, Ahmadi and Naghshineh [1] uses the *on-off* traffic model [3]. The *on* and *off* duration states have geometric distribution. During *on* state, there is cell generation at a constant rate and the *off* state is a silent period.

An *on-off* source can be characterized by its peak rate R, its load r and its mean burst length b. The switch model used by [1] is based initially on a single *on-off* source connected to a finite capacity queue (buffer) with constant service time, shown in Fig. 2.1.

Let $P_i(t, x)$ be the probability of the source be in state i when the occupancy of the buffer is x at the instant t.

If $i = 1 \Rightarrow$ source is active
If $i = 0 \Rightarrow$ source is idle.

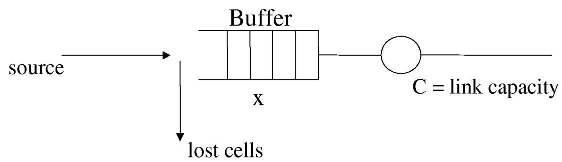

Figure 2.1: *ATM Multiplexer.*

Using the fluid flow model for the active periods of the source in the analysis of the buffer occupancy, the probability $P_i(t,x)$ can be found [7].

Supposing a buffer length K, the overflow probability ε (the required QoS) is given by [1]

$$\varepsilon = \beta \times \exp\left(- \frac{K(c - rR)}{b(1 - r)(R - c)c} \right) \quad (2.1)$$

where,

$$\beta = \frac{(c - rR) + \varepsilon r(R - c)}{(1 - r)c}.$$

The effective bandwidth, c, for a given source can be evaluated for a given ε.

The Eq. 2.1 has no explicit solution and can only be solved numerically. Thus, an approximation is suggested in [1].

Considering $\beta = 1$, the effective bandwidth can be evaluated and is given by [1]

$$\hat{c} = \frac{a - K + \sqrt{(a - K)^2 + 4Kar}}{2a} R \tag{2.2}$$

where

$$a = ln(1/\varepsilon)b(1\text{-}r)R.$$

For N sources, the effective bandwidth is given by

$$C_E = \sum_{i=1}^{N} \hat{c}_i \tag{2.3}$$

where $\hat{c}_i$ is the effective bandwidth of the source i.

Since the effect of the aggregation of the sources is not considered in Eq. 2.3, the bandwidth C_E can be overestimated. Thus, it is also considered an aggregated gaussian source.

For the gaussian source, the effective bandwidth is given by [1]

$$C_G = \rho + \alpha' \sigma$$

where,

$\alpha' = \sqrt{-2\ln(\varepsilon) - \ln(2\pi)}$,

$\rho = \sum_{i=1}^{N} \rho_i$, is the total average bit rate,

$\sigma = \sum_{i=1}^{N} \sigma_i$; σ_i is the standard deviation of the bit rate of the ith source.

The chosen bandwidth is given by

$$C_F = \min\{C_E, C_G\}.$$

If $C_F \le C_T$, where C_T is the total link capacity, the connection is accepted, otherwise it is rejected.

Kesidis, Walrand and Chang [2] proved the existence of effective bandwidth for a more general class of sources. The mathematical model is based on the assumption that the probability of the buffer occupancy obeys an exponential decay.

Furthermore, the model is based on the existence of the function

$$h(\delta) = \lim_{t \to \infty} \frac{\ln \mathrm{E}\{\exp(A(t)\delta)\}}{t}$$

where $A(t)$ is the number of generated cells during the time interval [0, t] and δ is a real value.

The effective bandwidth is given by [2]

$$c(\delta) = \frac{h(\delta)}{\delta}.$$

The evaluation of $h(\delta)$ is generally difficult. However, for the fluid flow markovian source (*on-off*) the equations above can be solved and the effective bandwidth is given by [12]

$$c = \alpha + \sqrt{\alpha^2 + \beta^2}$$

where,

$$\alpha = \frac{1}{2}\left(R - \frac{1}{\delta b} - \frac{1}{\delta T_{off}}\right) \quad \text{and} \quad \beta = \frac{R}{\delta T_{off}}.$$

T_{off} is the average time of the source silent periods,

$$T_{off} = \frac{b(1-r)}{r}$$

The value of δ is given by

$$\delta = \frac{-\ln \varepsilon}{K}$$

where ε is the required cell loss probability.

3. Simulation

In order to compare the performance of the two algorithms presented in Section 2, simulation was carried out. An ATM multiplexer handling the aggregated traffic of identical sources was used. The multiplexer is modeled as a finite buffer with capacity K, FIFO service discipline and one server with constant service rate. Two state Markov fluid sources (*on-off*) were used. These sources can be characterized by their peak rate R, average burst length b and load r. The simulation was developed using C programming language.

The link capacity was set at 155.52 Mbps and the required cell loss ratio, ε, was set at 10^{-3}. The traffic was applied to the buffer during 900 seconds which represents a considerable amount of traffic.

The simulation was divided into three groups, in order to analyze the algorithms behavior in relation to the variations of buffer capacity, average burst length and source load.

An additional simulation was executed in each of the three groups to determine upper bound of connections acceptance. Through successive adjusts, the number of accepted connections was increased or decreased until the obtained cell loss ratio was as close as possible, but lower than its required value. The results represent the maximum number of connections can be handled by the multiplexer for the specified CLR. The algorithms behavior is compared to this upper bound.

4. Results

For the performance study of the algorithms in relation to the variation of buffer length, the following source parameters were used: peak rate = 8 Mbps, average burst length = 100 cells and load = 0.2. Low load was used in order to explore the statistical gain.

It can be seen in Fig. 4.1 that the number of accepted connections by the two algorithms is very different in small buffer region (K < 1000 cells). The GAN (Guérin, Ahmadi and Naghshineh) algorithm [1] shows better link utilization. The gaussian approximation permits the GAN algorithm to reach a better result than that obtained by Kesidis, Walrand and Chang (KWC) algorithm when buffer is small. Increases in the buffer capacity beyond 5000 cells have little effect. Both schemes get closer to the upper bound limit for large buffer capacity.

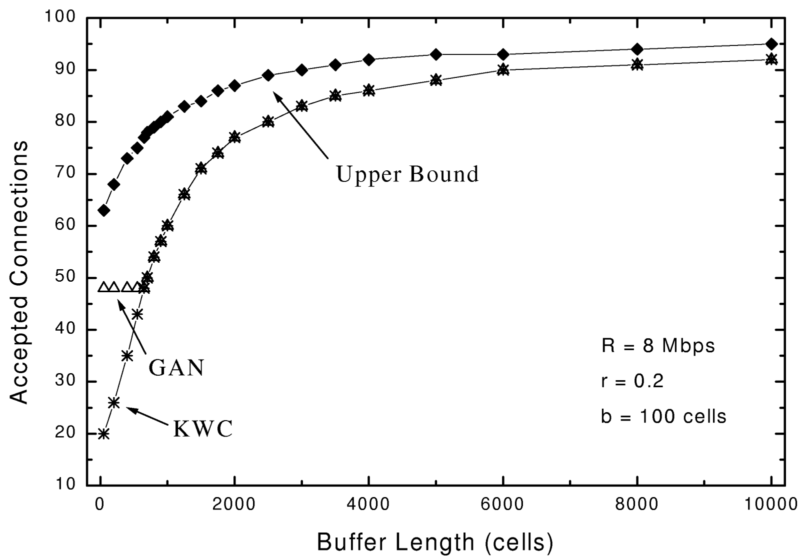

Figure 4.1: Number of accepted connections as a function of buffer length.

Fig. 4.2 shows that the link utilization is high only for large buffers. With buffer capacity below 1000 cells the link utilization is low. A buffer capacity of 2000 cells is enough to reach a link utilization of approximately 0.8.

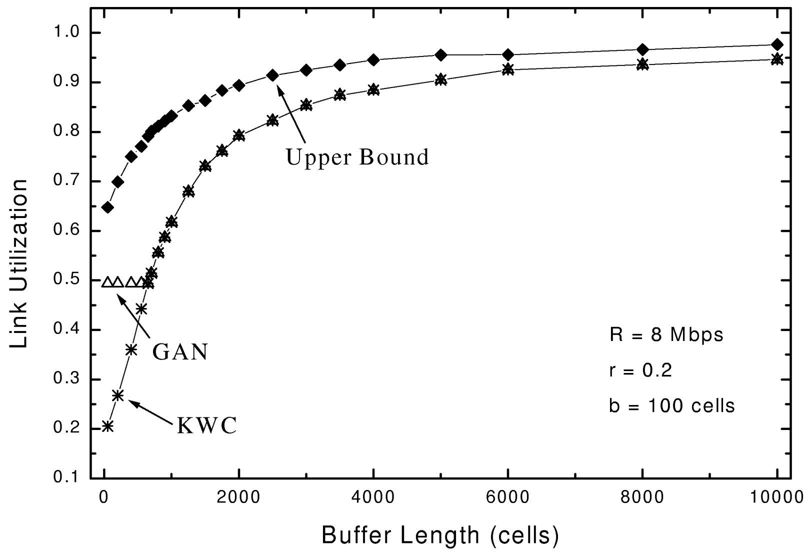

Figure 4.2: Link utilization as a function of buffer length.

The average waiting time of cells in buffer is an important parameter because real time applications cannot bear large cell delays. Large buffer allows high link utilization, although there is a commitment between utilization and cell delay. Fig. 4.3 shows that for small buffers (K < 2000 cells) a short cell delay is obtained and large buffer capacity implies long waiting time.

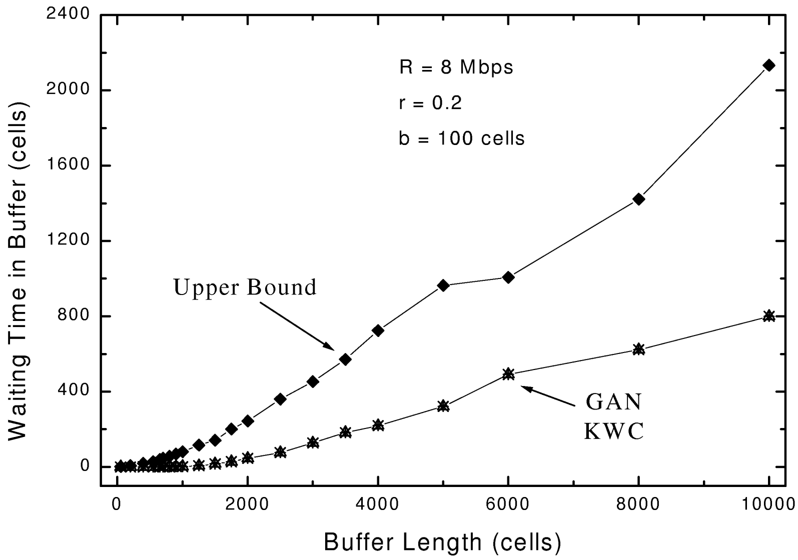

Figure 4.3: Average waiting time in buffer as a function of buffer length.

For the performance study of the algorithms in relation to the burst size the sources have peak rate of 8 Mbps, buffer capacity of 4000 cells and load of 0.3. Once more, we can see that the GAN algorithm [1] achieves the best results, as is shown in Fig. 4.4. The two algorithms have significant differences when burst length is large. For small burst length there is no difference in the number of accepted connections.

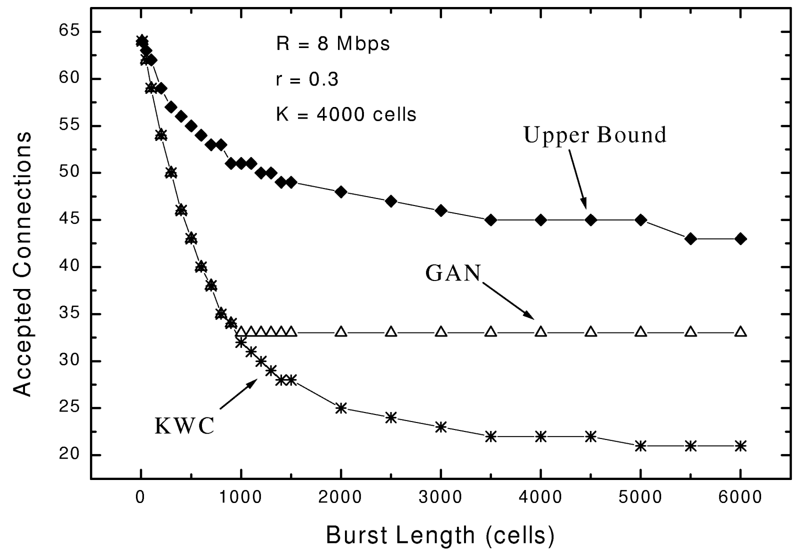

Figure 4.4: Number of accepted connections as a function of burst length.

The link utilization decays rapidly as the average burst length increases (Fig. 4.5), but it can be very high for burst lengths shorter than 500 cells. Obviously, this rapid decay is influenced by the buffer capacity. Larger values would permit a slower decay.

Again, the average waiting time is long when link utilization is high (Fig. 4.6). If waiting time is the most restrictive parameter, it is necessary to operate at lower utilization levels.

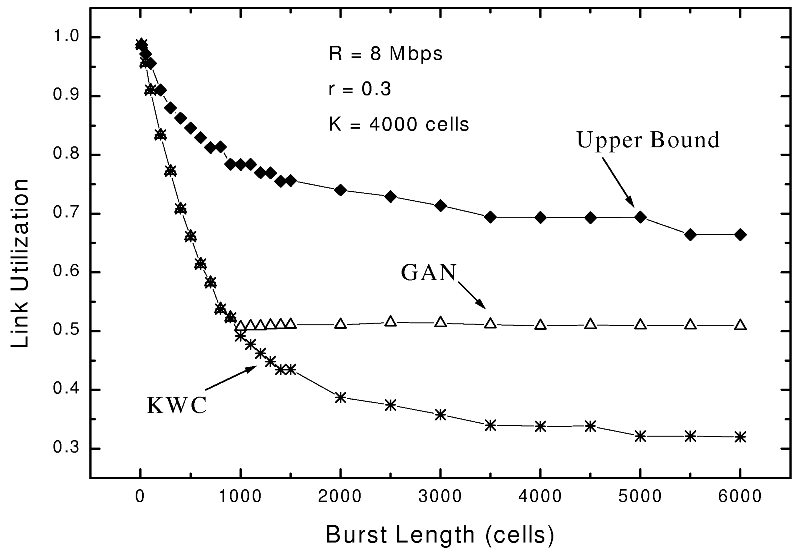

Figure 4.5: Link utilization as a function of burst length.

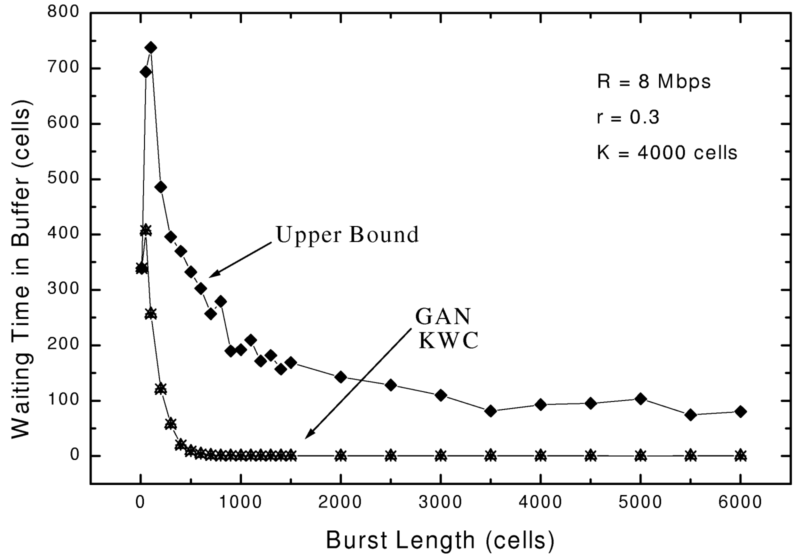

Figure 4.6: Average waiting time in buffer as a function of burst length.

For the performance study of the algorithms in relation to the load of sources, it was used peak rate of 8 Mbps, buffer capacity of 1000 cells and average burst length of 200 cells. The results show that both algorithms are conservative. The KWC (Kesidis, Walrand and Chang) algorithm [2] is highly conservative when the source load is low as is shown in Fig. 4.7. Additional simulations showed that this difference between the algorithms diminishes when the K/b ratio is bigger.

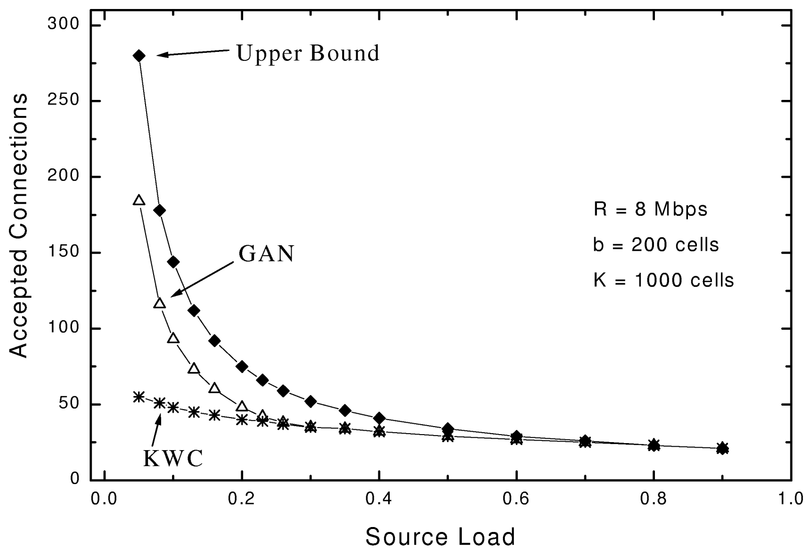

Figure 4.7: Number of accepted connections as a function of source load.

In order to obtain a high link utilization when sources have low load, a sufficiently large buffer is necessary in order to store possible long sequences of aggregated bursts. The buffer length used had not sufficient capacity (Fig. 4.8). However, this limits the cell waiting time at low values (Fig. 4.9).

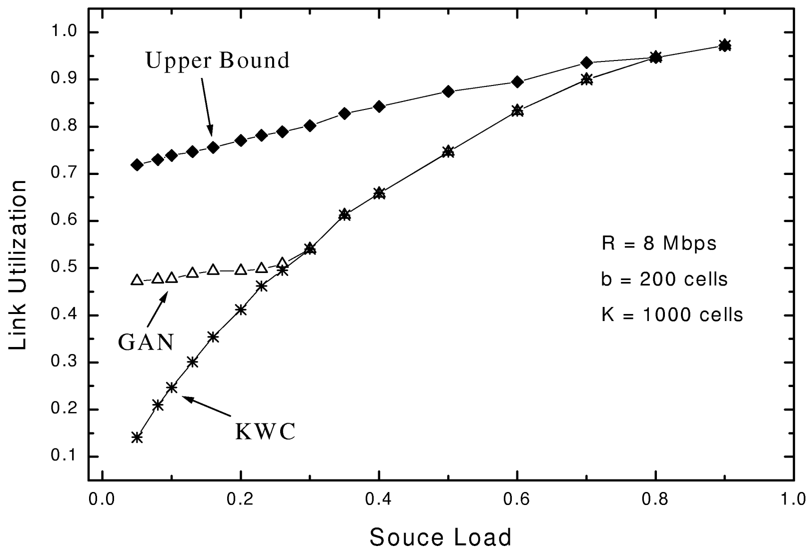

Figure 4.8: Link utilization as a function of source load.

The average waiting time decay observed in Fig. 4.9 for load values above 0.7 can be explained as follow. When load is equal to 0.7, 26 connections are accepted and the resulting cell loss ratio is close to the required CLR. Load value of 0.8 results in the acceptance of only 23 connections, with cell loss ratio equal to zero. However, just one more connection is sufficient to cause CLR violation. Thus, larger loads cause a lower statistical gain and the connection acceptance is more influenced by link rate rather than buffer length, which tends to show lower occupation and cell waiting time.

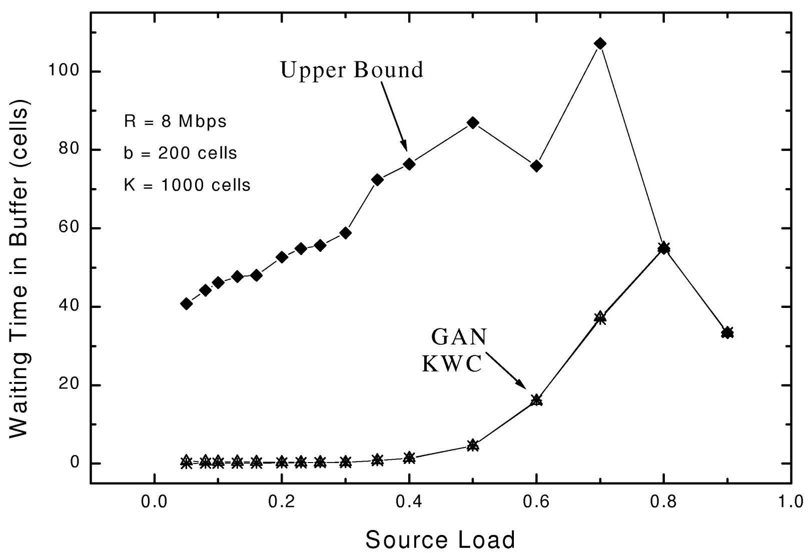

Figure 4.9: Average waiting time in buffer as a function of source load.

5. Conclusions

The performance study and comparison of two CAC algorithms were carried out in this paper. The behavior of the algorithms in relation to the buffer lengths, burst size and source load was investigated.

The GAN algorithm showed better performance for shorter buffer lengths. For larger buffer a similar performance is obtained for both algorithms.

The burst length have strong effect on algorithms behavior. As the burst length increases the number of connections decreases rapidly, because the effective bandwidths of sources are overestimated. The algorithms are more conservative when the relation K/b is smaller.

For small load values ($r < 0.2$), the KWC algorithm performance is not satisfactory. The gaussian approximation used by GAN algorithm permits a better performance, although still being conservative. However, the performance of both algorithms tends to approximate when the K/b relation is increased.

6. References

[1] R. Guérin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," IEEE J. Select. Areas Commun., vol. 9, pp. 968-981, 1991.

[2] G. Kesidis, J. Walrand, and C.-S. Chang, "Effective bandwidths for multiclass Markov fluids and other ATM sources," IEEE Trans. Networking, vol. 1, no. 4, pp. 424-28, Aug. 1993.

[3] H. Michiel, and K. Laevens, "Teletraffic engineering in a broad-band era," Proceedings of the IEEE, vol. 85, no. 12, pp. 2007-2033, Dec. 1997.

[4] H. Saito, and K. Shiomoto, "Dynamic call admission control in ATM Networks," IEEE J. Select. Areas Commun., vol. 9, no. 7, pp. 982-989, 1991.

[5] A. W. Berger, and W. Whitt, "Effective Bandwidths with priorities," IEEE/ACM Transactions on Networking, vol. 6, no. 4, pp. 447-460, 1998.

[6] H. Saito, "Call admission control in an ATM network using upperbound of cell loss probability," IEEE Trans. Commun., vol. 40, pp. 1512-1521, 1992.

[7] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," Bell Sys. Tech. J., vol. 61, pp. 1871-1894, 1982.

[8] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand, and R. Weber, "Admission control and routing in ATM networks using inferences from measured buffer occupancy," IEEE Trans. Comm., Feb/Mar/Apr, 1995.

[9] I. Hsu, and J. Walrand, "Dynamic bandwidth allocation for ATM switches," J. Appl. Prob., 33, pp. 758-771, 1996.

[10] R. Bolla, F. Davoli, and M. Marchese, "Bandwidth allocation and admission control in ATM networks with service separation," IEEE Comm. Magazine, pp. 130-137, May, 1997.

[11] J.-Y. Le Boudec, and R. Nagarajan, "A CAC algorithm for VBR connections over a VBR trunk," Proceedings of 15th International Teletraffic Congress – ITC, vol. 1, pp. 59-70, Washington, DC, USA, 1997.

[12] J. C. Oliveira, I. S. Bonatti, P. L. D. Peres and A. K. Budri, "Cell level performance approach for link dimensioning of ATM networks," Proceedings of ITS 98, vol.1, pp. 189-194, São Paulo, SP, Brazil, 1998.