

# Benefit of Admission Control in Aggregation Network Dimensioning for Video Services

Gert Van Hoey<sup>1</sup>, Danny De Vleeschauwer<sup>1</sup>,  
Bart Steyaert<sup>2</sup>, Véronique Ingelbrecht<sup>2</sup> and Herwig Bruneel<sup>2</sup>

<sup>1</sup> Alcatel Bell, R&I/NSG, Fr. Wellesplein 1,  
B-2018 Antwerp, Belgium

{Gert.Van\_Hoey, Danny.De\_Vleeschauwer}@alcatel.be

<sup>2</sup> University of Ghent, SMACS, Sint-Pietersnieuwstraat 41,  
B-9000 Ghent, Belgium  
{bs, vi, hb}@telin.ugent.be

**Abstract.** We derive a methodology to determine the effect of per service session admission control on the resource usage in an aggregation network for the delivery of video services. Given the network capacity and the characteristics of the service, we determine the number of subscribers that can be served in the scenarios with and without admission control. The results depend on the activity grade of the subscribers, and on the session blocking probability and/or the packet loss probability that can be tolerated. For a broad range of both probability parameters, the comparison shows a clear benefit of admission control, which permits an increase of the number of subscribers by up to 30%. The benefit of admission control decreases as the subscriber activity grade increases.<sup>(a)</sup>

## 1 Introduction

An important problem encountered by network access providers (NAP) is the dimensioning of the aggregation network through which the NAP's customers can connect to Internet Service Providers (ISPs) or Application Service Providers (ASPs). An example is a DSL aggregation network, in which DSL access multiplexers (DSLAMs) aggregate traffic from typically hundreds of DSL subscribers. The traffic from the DSLAMs is then further aggregated by ATM- or Ethernet-based switched networks.

Dimensioning deals with determining the relationship between the number of customers attached to the aggregation network, the traffic characteristics of the services and applications used by these customers, and the network resources (link bandwidths, switch capacities) that are available or required in the network.

---

<sup>(a)</sup> This work was carried out within the framework of the project ARGUS sponsored by the Flemish Institute for the promotion of Scientific and Technological Research in the Industry (IWT).

A straightforward approach for determining the required network resources is overdimensioning, which means that the network is dimensioned such that it can deal with the worst-case scenario that all customers are simultaneously on-line and using the service (voice-over-IP, video-on-demand (VoD), ...). In general, overdimensioning is not feasible because of limited aggregation network resources, which is particularly true for services requiring a relatively high bandwidth (e.g. video).

As a consequence, we need to take into account the fact that not all customers are consuming network resources simultaneously, and dimension such that the available resources, while considerably lower than in the overdimensioning case, still suffice to ensure users are fully satisfied. The user satisfaction is mainly determined by two factors. The first one is the availability of the subscribed service, which is the ability to initiate a service session at any time the customer wishes to do so. The second one is the quality of the transmitted information, which is determined by the extent to which packet delay and packet loss restrictions are met during an ongoing service session.

The dimensioning of the aggregation network will depend on whether or not the NAP performs admission control. Admission control (or call acceptance control, CAC) limits the number of simultaneously ongoing sessions (e.g. video-on-demand sessions) to a predetermined number, and thus, if needed, prevents a user from setting up a session at a moment when the maximum number of sessions is temporarily reached. The advantage of applying admission control is that it enables the NAP to guarantee the quality of transmitted information (packet loss, delay) because the maximum number of simultaneous sessions is known beforehand. The disadvantage is that from time to time, a customer has no access to the subscribed service (less than 100% availability). In a scenario without admission control, the availability of the service is 100%, but the possibility exists that temporarily, the amount of required network resources exceeds the provisioned resources, resulting in packet loss and excessive delay, and thus reduced service quality for all customers with ongoing sessions.

This paper compares the scenarios with and without CAC in terms of the network resources that need to be provisioned for a given customer population and service characteristics (or, equivalently, in terms of the number of subscribers that can be connected to the network, given the network resources). We focus on video services, which play a prominent role in emerging network-based applications and services. Because of the relatively high bandwidth requirements, the absolute differences in terms of required network capacity are expected to be substantial, and thus very relevant to network operators with limited resources in their aggregation network.

The remainder of the paper is organized as follows. Section 2 introduces the problem that is considered and describes the relevant parameters of the network, the users, and the service. The models on which this methodology relies, are introduced in section 3. Section 4 describes the methodology that is followed to determine the maximum number of VoD subscribers in the scenarios with and without CAC, while section 5 presents and discusses the results of the comparison between the two scenarios. Finally, section 6 concludes this paper.

## 2 Problem statement

Figure 1 shows the considered aggregation network and the various network and traffic parameters that play a role in the dimensioning study of this paper. It shows a DSL aggregation network in which a group of  $N$  DSL subscribers are attached to the same DSLAM. The DSLAM's so-called "uplink" connects it with a switch in the aggregation network, that aggregates the traffic of several DSLAMs. Typical uplink capacities are 599 Mbps (STM-4, with 622 Mbps SDH bit rate) in the case of an ATM aggregation network, and 1 Gbps in the case of Ethernet. It should be noted that the methodology and results in this paper are equally applicable to aggregation networks other than DSL.

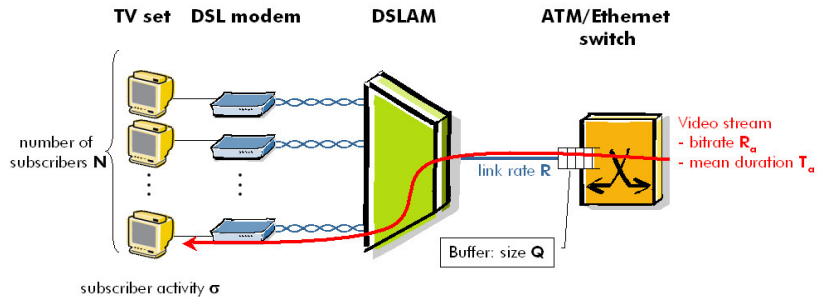


Fig. 1. Network setup

We assume that if a capacity bottleneck exists in the aggregation network, it is located in the aggregation segment (segment where resources are shared by multiple customers) closest to the customer side. In the DSL network of figure 1, this is the link between the DSLAM and the switch to which it is connected (the link modem-DSLAM is dedicated to a single subscriber, and hence not shared with other subscribers). In cable networks, the potential bottleneck would be located in the cable segment shared by a group of subscribers. Because the VoD service is considered in this paper, it is the network capacity in the downstream direction (from network towards user) that potentially forms a bottleneck. This means that the downstream link rate  $R$  and the size  $Q$  of the queue in the switch feeding the link towards the DSLAM, are the two parameters that determine to what extent packets of video streams can be lost. The packet loss probability will be denoted by  $P_{packet}$ .

Video streams are assumed to consist of 1500 byte packets and all have the same constant bit-rate (CBR), denoted by  $R_a$  (Mbps). The various video streams may have different durations, the average duration being denoted by  $T_a$  (sec). The parameter  $\sigma$ , the subscriber activity grade, indicates the probability that an individual VoD subscriber is actually watching a video stream at a randomly chosen moment in time.

In this paper, we take the point of view of a network operator that wants to determine the maximum number of VoD subscribers  $N$  that can be supported per DSLAM in his existing aggregation network. This means that we assume the network parameters  $R$  and  $Q$  as given, and that based on the values of  $R_a$ ,  $T_a$ , and  $\sigma$ , the maxi-

mum  $N$  is determined that still guarantees a packet loss ratio below  $P_{packet}$ , and — in the case of admission control — a session blocking probability lower than  $P_{session}$ .

We confine ourselves to a network scenario with homogeneous (i.e., identical) sources, because a number of results in the literature suggest ([6]) that this serves as a worst-case to a heterogeneous scenario with comparable mean and peak rate. Hence, although many realistic traffic scenarios will include heterogeneous VoD sources, the homogeneous case can be used for dimensioning purposes.

### 3 Models for VoD performance assessment

#### 3.1 Models to assess the packet loss

The VoD quality will rapidly degrade when the packet loss becomes too high. It is well known (e.g. [5,7]) that under fairly general conditions the packet loss in a finite-capacity single-server queue of size  $Q$  can be accurately approximated by the probability that the queue length (represented here by the random variable  $u$ ) in an infinite capacity exceeds  $K$ . Therefore, in the following sections we focus on the calculation of the  $P$ -quantile of the queue length in an infinite-capacity queue, defined as

$$Q_u(P) \triangleq \min_Q \{ \Pr[u > Q] < P \} \quad (1)$$

This quantity calculated from the  $N \times \text{Geo}/G/1$  and  $N \times \text{MMBP}/G/1$  queueing model will be used to estimate the buffer size corresponding to a packet loss  $P$  in the associated finite-capacity queue.

When a subscriber is active (i.e., a session is ongoing), packets will be generated at a rate  $\gamma$  corresponding to  $R_a$ . The packet length (expressed in number of slots) is assumed to be generally distributed according to the probability generating function (pgf)  $S(z)$ . Let  $L_d$  (bytes) be the basic unit we use for expressing the length of the data packets and let  $S_l$  correspond to the slot length, then we clearly have that

$$S_l = 8L_d / R \quad ; \quad \gamma = S'(1)R / R_a \quad (2)$$

In a number of traffic scenarios it suffices to consider constant-length packets, in which case  $S(z)=z$ .

##### 3.1.1 The $N \times \text{Geo}/G/1$ model

Let us consider the situation where  $N$  subscribers are active, each receiving packets at a rate  $\gamma$ . Since VoD sessions are typically very long (i.e., hundreds of seconds or more), it is reasonable to model the system performance under these circumstances by the steady-state behaviour of the discrete-time  $N \times \text{Geo}/G/1$  queue. If we define the queue length  $u$  as the amount of work in the queue (not including the packet being transmitted, if any), then it can be deduced that its steady-state pgf is given by ([1])

$$U(z) = (1-\rho) \left\{ \frac{x-1}{x-S(A(x))} \right\}_{x=S(z)}, \quad (3)$$

where  $\rho=N\gamma$  and  $A(z)$  is the pgf describing the number of packet arrivals per slot and is therefore given by  $A(z)=(1-\gamma+\gamma z)^N$ .

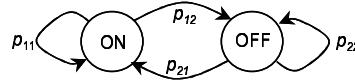
From (1) it follows that we need an efficient way to calculate the tail distribution of  $u$ ; in [4] it is shown that this quantity can be accurately approximated by using a dominant-pole approximation for  $U(z)$ , leading to

$$\Pr[u > Q] = \frac{-C}{z_0 - 1} z_0^{-Q-1}, \quad (4)$$

where  $z_0$  is the smallest pole of  $U(z)$  and  $C$  the corresponding residue. In view of (3), it becomes clear that  $z_0$  is the (smallest) real zero of  $z=A(S(z))$  satisfying  $z_0 > 1$  (we refer to [4] for further detail concerning these results).

Whenever the value of  $N$  becomes sufficiently high,  $A(z)$  comes close to describing a Poisson arrival process, and the results of the M/G/1 model can be applied instead. This property will be exploited in the next sections.

### 3.1.2 The $N \times$ MMBP/G/1 model.



**Fig. 2.** The ON/OFF subscriber model

Subscribers will not be watching VoD all the time, but will switch between active (during which a session is ongoing) and non-active periods. A popular approach in this case is to model each subscriber by a two-state ON/OFF source model in order to capture this type of behaviour. The ON state obviously corresponds to an ongoing VoD session during which packets are generated at a rate  $\gamma$ , while no packets are received during an OFF period. A graphic description of this subscriber conduct, in addition to the corresponding transition parameters, is shown in Fig. 2. The subscriber activity  $\sigma$  then equals  $p_{21}/(p_{21} + p_{12})$ , and the transition probabilities can be derived from the VoD source characteristics in the following manner :

$$\begin{aligned} p_{12} &= S_l / T_a \quad ; \quad p_{11} = 1 - p_{12} \\ p_{21} &= \frac{1-\sigma}{\sigma} p_{12} \quad ; \quad p_{22} = 1 - p_{21} \end{aligned} \quad (5)$$

An important quantity in our source description is the  $2 \times 2$  matrix  $\mathbf{Q}(z)$ , defined as

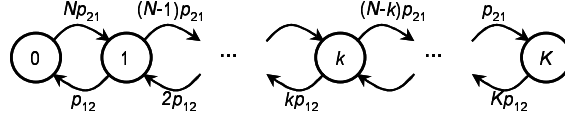
$$\mathbf{Q}(z) = \begin{bmatrix} p_{11}(1-\gamma+\gamma z) & p_{12} \\ p_{21}(1-\gamma+\gamma z) & p_{22} \end{bmatrix}, \quad (6)$$

which fully captures the ON/OFF subscriber behaviour.

The packet arrival process described above is of the Markov Modulated Bernoulli Process (MMBP) type, and the analysis of an infinite-capacity single queue fed by this type of sources can be carried out by resorting to a pgf-approach ([8]). This leads to the following expression for the steady-state pgf  $U(z)$  of the queue length  $u$  :

$$U(z) = (1-\rho) \sum_{m=0}^N \frac{(x-1)}{x - S(\lambda_1(x)^m \lambda_2(x)^{N-m})} \xi_m(x) \Big|_{x=S(z)}. \quad (7)$$

In this expression,  $\lambda_j(z)$ ,  $1 \leq j \leq 2$ , represent the eigenvalues of  $\mathbf{Q}(z)$ . The functions  $\xi_m(z)$  contain  $N+1$  boundary probabilities that can be calculated by exploiting the



**Fig. 3.** The  $N \times \text{Geo}/\text{Geo}/K/K$  state transition diagram

analyticity of  $U(z)$  inside the complex unit disk, leading to a set of  $N+1$  linear equations that needs to be solved. However, in order to keep the calculations as efficient as possible we prefer to use an approximation that has been shown to be sufficiently accurate, which yields

$$\xi_m(z) \cong \frac{N!}{m!(N-m)!} \left( \frac{\bar{\sigma}^T \mathbf{Q}(0) \bar{u}_1(z)}{\bar{\sigma}^T \mathbf{Q}(0) \bar{1}} \right)^m \left( \frac{\bar{\sigma}^T \mathbf{Q}(0) \bar{u}_2(z)}{\bar{\sigma}^T \mathbf{Q}(0) \bar{1}} \right)^{N-m}, \quad (8)$$

with  $\bar{\sigma} \triangleq [\sigma \quad 1 - \sigma]^T$ ,  $\bar{1} \triangleq [1 \quad 1]^T$ , and  $\bar{u}_j(z)$  represent the column eigenvectors of  $\mathbf{Q}(z)$  associated with  $\lambda_j(z)$  satisfying  $\bar{u}_1(1) + \bar{u}_2(1) = \bar{1}$ . This approximation is based on the observation that the event of the buffer being empty at the beginning of a slot, implies that there were no packet arrivals during the preceding slot.

Once again, the tail distribution of the queue length is calculated by adopting a dominant-pole approximation. In this case however, it does not suffice to use an expression similar to (4) since a thorough evaluation of this approach has revealed that multiple poles of  $U(z)$  can now contribute significantly to a tail probability. Hence we will use the following formula for tail probability calculations :

$$\Pr[u > Q] \cong \sum_{m \leq N} \frac{-C_m}{z_{0,m} - 1} z_{0,m}^{-Q-1} \quad (9)$$

where the pole  $z_{0,m}$  of  $U(z)$  is the real solution of  $z = \lambda_1(S(z))^m \lambda_2(S(z))^{N-m}$  satisfying  $z_{0,m} > 1$ , and  $C_m$  its corresponding residue. Starting from  $m=N$ , decreasing values

of  $m$  correspond to ‘less dominant’ poles, and for any given value of  $Q$  subsequent terms in (9) should be calculated, until the value of the term calculated last is negligible compared to the sum of the ones calculated before. For a wide range of parameter values, it has been found that this approach yields an efficient procedure for the calculation of queue length quantiles.

### 3.2 Models to assess the session loss : the $N \times \text{Geo}/\text{Geo}/K/K$ queueing system

When CAC is applied, we need some means to calculate the session blocking probability, for a population of  $N$  subscribers of which  $N_{stream}$  are allowed to be active simultaneously. The calculation of blocking probabilities is usually done by considering an appropriate server-loss (bufferless) model. Therefore, in view of the queueing model presented in section 3.1.2 for evaluation of the packet loss, the natural way to proceed is to calculate session blocking probabilities by means of a discrete-time  $N \times \text{Geo}/\text{Geo}/K/K$  (where  $K \equiv N_{stream}$  in our case) queueing model. In order to simplify the analysis, let us furthermore assume that the probability of two or more events (an event being an active subscriber terminating his session or a passive subscriber attempting to initiate a new session) occurring during a slot is negligible compared to the probability of a single event occurring. This assumption will yield accurate results when sessions (and silent periods between sessions) are typically very long, which is clearly the case for the VoD scenario under study. Under these circumstances, Fig. 3 represents the state transition diagram from which the steady-state probabilities of being in state  $k$  can be readily derived. We must point out that this diagram also describes the behaviour of the continuous-time  $M/M/K/K/N$  queueing system, also known as the Engset loss system ([1]).

The session blocking probability in the  $N \times \text{Geo}/\text{Geo}/K/K$  model is therefore given by the Engset blocking probability. If we want to calculate the maximum number of subscribers that we can accommodate under the restriction that the session blocking probability may not exceed  $P_{session}$ , we must solve the following equation

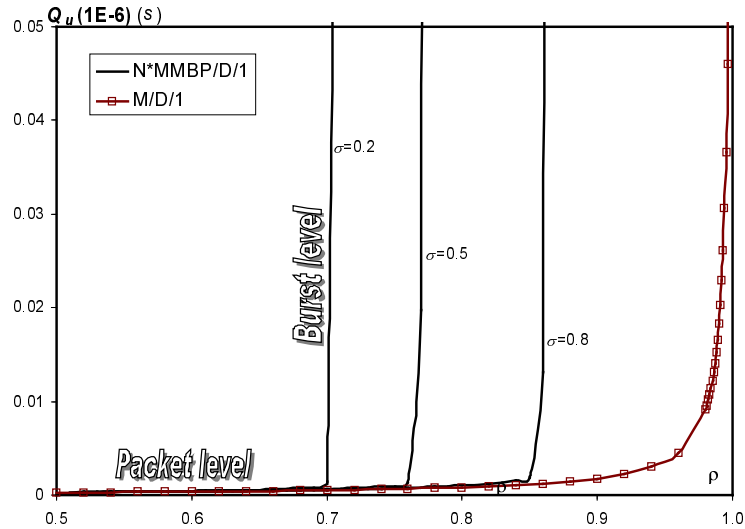
$$\max_N \left\{ \frac{C_{N-1}^K \sigma^K (1-\sigma)^{N-K-1}}{\sum_{j=0}^K C_{N-1}^j \sigma^j (1-\sigma)^{N-j-1}} \right\} \leq P_{session} \quad (10)$$

i.e., we must calculate the largest value of  $N$  for which the inequality still holds.

### 3.3 A numerical example: packet versus burst level behaviour

Let us illustrate the typical buffer behaviour for  $R=450$  Mb/s,  $R_a=2$  Mb/s and  $T_a=5400$  sec, and for constant-length VoD packets of 1500 byte (i.e.,  $L_d=1500$ ). In figure 4, we have plotted the  $10^{-6}$  quantile of the queue length, obtained from the  $N \times \text{MMBP}/D/1$  model, versus the load  $\rho$  for various values of the subscriber activity

$\sigma$ . In this figure we express the quantile  $Q_u(P)$  in seconds, because this immediately relates to the maximum delay that can be expected by a newly arriving packet.



**Fig. 4.** Packet versus burst level buffer behaviour

Here we already observe the typical quantile behaviour that captures the essence of the VoD buffer performance. We can roughly distinguish two regions for the load  $\rho$ . On the one hand, an area of relatively small loads where the quantiles remain stable for increasing values of the load. This means that in this area packet loss can be avoided by assigning sufficient buffer space to the VoD buffer; the required buffer space however remains relatively small (i.e.,  $\sim 5$  ms), and the buffer performance can be regarded upon as being dominated by the *packet level* behaviour. On the other hand, for further increasing values of the load, one reaches a point where the quantiles suddenly explode : small increments of the load lead to huge increases of the quantiles. This point can be considered as the turning point where the buffer performance commences to be dominated by the *burst level* behaviour. In this region, large amounts of work can accumulate in the buffer, and whenever this occurs, the buffer will not be able to accommodate a newly arriving session. Note that, for these parameter values, a buffer of  $0.01s$  can contain approximately 375 packets, while a VoD subscriber which is active during just 1 minute generates 10 000 packets on average. This explains why trying to buffer entire bursts becomes futile, especially if one wants to guarantee sufficiently small packet delays (and delay jitter), as well as small packet loss, as a measure of QoS.

When the packet level behaviour dominates the buffer performance, it suffices to have relatively small buffers in the switch-DSLAM (which also guarantees small packet delays) and the exact value of the buffer size  $Q$  is not important as long as it is ‘high enough’, which we assume to be the case in the remainder of this paper.



## 4 Methodology

### 4.1 Without admission control

In the scenario without admission control, the NxMMBP/G/1 model is applied to describe the switch-DSLAM buffer behaviour. Clearly, in view of the discussion in Section 3.3, from a network operator's point-of-view, the targeted working area is the range of  $\rho$  values where the packet level dominates, and the 'turning point' determines the maximum value of  $\rho$  that can be supported when no CAC is applied. For the given values of  $R$  (450 Mbps) and  $P_{packet}$  ( $10^{-6}$ ) and the traffic parameters  $R_a$  (2 Mbps),  $T_a$  (5400 sec) and  $\sigma$  (0.2), the turning point is located at  $\rho=0.709$ . (cfr.  $Q_u(P_{packet})$  in Fig. 4) Based on this value of  $\rho$ , the maximum number of subscribers  $N$  is then determined as  $\lfloor R\rho/(\sigma R_a) \rfloor$ , which yields  $N_{noCAC}=797$ .

### 4.2 With admission control

In the case of admission control, the maximum number of subscribers  $N$  is determined in two steps. First, the M/G/1 model is used to determine the maximum load  $\rho$  that can be tolerated on the switch-DSLAM link. In Fig. 4 we have plotted the  $10^6$  quantile of the M/D/1 queue, from which we can deduce that for  $P_{packet}=10^{-6}$  and if we allow packet delays up to 0.05s, then the maximum tolerable load is nearly 1 (i.e.  $\rho=0.996$ ), leading to  $N_{stream}=\lfloor R\rho/R_a \rfloor=224$  (note that in most cases the value of  $\rho$  will be close to 1 leading to  $N_{stream}\approx\lfloor R/R_a \rfloor$ ). Secondly, this value of  $N_{stream}$  is plugged into the N×Geo/Geo/K/K model (i.e., equation (10)) to determine the maximum number of subscribers that can be sustained for a given activity rate  $\sigma$  and blocking probability  $P_{session}$  when CAC is applied. For  $P_{session}=10^{-4}$ , this model leads to  $N_{CAC}=915$ . This means that thanks to CAC, 915 subscribers can be supported instead of 797 in the case without CAC, which is an increase by 14.8%.

## 5 Results and discussion

In this section, we quantify the number of VoD subscribers  $N$  that can be supported by an aggregation network, in the scenarios with ( $N_{CAC}$ ) and without ( $N_{noCAC}$ ) admission control, using the models and methodology described in the previous sections.

A video session is assumed to require a video codec  $R_a = 2$  Mb/s and last on average  $T_a = 5400$  s. We consider various values for the activity grade  $\sigma$ , (0.2, 0.5 and 0.8). A video packet has a size of 1500 byte. The aggregate bit rate that is available for the video service amounts to  $R = 450$  Mb/s. This could, e.g., correspond with a situation with a 622 Mbps uplink on which a capacity of 172 Mbps needs to be available at any time for services other than VoD.

For the performance parameters we consider various values. For the tolerated packet loss  $P_{packet}$  we consider values from  $10^{-10}$  to  $10^{-3}$ , whereas for the session loss probability  $P_{session}$  (or blocking probability) we consider values from  $10^{-8}$  to  $10^{-3}$ . The values that are tolerated depend on the type of service the user is requesting.

Video codecs are not very robust against packet loss (a value of the order of  $10^{-6}$  is just bearable for MPEG codecs [3]). Moreover, packet loss may be very bursty: certainly in the NxMMBP/G/1 model there are indications that the system will slowly alternate from an under-load state in which no packets are lost at all to an overload state where the packet loss is very high (e.g. Fig. 4). Even though the overall (averaged) packet loss (the performance parameter that is resulting from the model studied) is low enough, it still could be that the loss during the (long-lasting) overload state would be detrimental. Therefore, the tolerated packet loss probability  $P_{packet}$  should be chosen very conservatively for video.

The choice of the tolerable session loss probability  $P_{session}$  is set by the operator. A value of  $10^{-3}$  seems reasonable, but the operator may choose a lower value.

In order to compare the scenario with AC with the one without AC, we define the gain due to AC as:

$$\frac{N_{CAC} - N_{noCAC}}{N_{noCAC}} \quad (11)$$

Table 1 gives the gain for various values of the tolerated packet loss and session loss for users that are only active during 20 % of the time on average. A considerable gain (in terms of the number of subscribers that can be supported) can be obtained by implementing an admission control mechanism, if we choose the tolerated  $P_{packet}$  conservatively (as we need to do for a video service) and as long as the session loss probability  $P_{session}$  is not chosen too small. E.g., if we require the packet loss ratio to be below  $10^{-8}$ , and put forward  $10^{-4}$  as an acceptable blocking probability for the VoD service, the scenario with CAC enables the operator to have 20.7% more VoD subscribers attached to a single DSLAM than in the case without CAC. From Table 2 and Table 3 it can be seen that the benefit from implementing an admission control mechanism becomes gradually smaller as the activity grade of the user increases. The tables 1-3 further point out that for services that may be more tolerant w.r.t. packet loss and less w.r.t. blocking, the CAC case can turn out to enable less subscribers than the case without CAC.

In order to put the results of tables 1-3 in perspective, two important remarks should be made concerning the interpretation of the packet loss probability in the scenarios with and without CAC. First, in the case of CAC,  $P_{packet}$  has to be interpreted as the packet loss probability during time periods where the maximum number of VoD sessions ( $N_{stream}$ ) are going on simultaneously. Because these time periods only represent a fraction of the total time,  $P_{packet}$  should be regarded as a "worst-case" packet loss ratio, and an "average" packet loss probability would have a smaller value. Secondly, in the case without CAC, packet loss is expected to be bursty. During long time intervals where the number of active users does not exceed  $\lfloor R/R_a \rfloor$ , the packet loss ratio will be very low. From time to time, however, the number of active users will exceed  $\lfloor R/R_a \rfloor$ , giving rise to a large packet loss ratio of approximately  $1/\lfloor R/R_a \rfloor$ . The time periods with high packet loss must be expected to last in the order of seconds or tens of seconds, given the fact that video streams of long duration are considered. This means that even if a low value of the "average"  $P_{packet}$  is put forward, time periods with high packet loss can occur in the case without CAC.

From these considerations, it is clear that if models were available to quantify the discussed issues, the comparison between the scenarios with and without CAC would turn out more advantageous (or less negative) for CAC than currently indicated by the tables 1-3.

Another consideration to take into is the effect of an increase in the value of the subscriber activity grade  $\sigma$ . In practice, the value of  $\sigma$  will be determined based on observations of the average number of subscribers that is simultaneously active during the busy hour of the service. It is possible, however, that at a certain moment in time, the demand for the service is significantly larger (e.g. caused by an unexpected important news item), such that the corresponding  $\sigma$  is larger than the value taken into account for dimensioning. The important distinction is that on the one hand, in the scenario with CAC, the higher  $\sigma$  will lead to an increased blocking probability  $P_{session}$ , while the packet loss guarantee still holds for all the service sessions that are in progress. On the other hand, in the scenario without CAC, the higher  $\sigma$  leads to a larger number of ongoing sessions, with all ongoing sessions experiencing increased packet losses. This increase in packet loss could be so severe that it renders the service practically useless for all ongoing sessions. It is difficult to point out which one of both effects, high blocking probability or excessive packet loss, is most undesirable. Still, a strong argument in favour of CAC is that it guarantees that a certain number ( $N_{stream}$ ) of service sessions can always take place with the necessary packet loss guarantees.

**Table 1.** The gain due to the activation of an AC mechanism for an activity grade  $\sigma$  of 0.2

		$P_{session}$					
		1.E-08	1.E-07	1.E-06	1.E-05	1.E-04	1.E-03
$P_{packet}$	1.E-03	-8.1%	-5.5%	-2.5%	0.8%	4.8%	10.0%
	1.E-04	-5.0%	-2.3%	0.8%	4.3%	8.4%	13.7%
	1.E-05	-2.1%	0.7%	3.9%	7.4%	11.7%	17.2%
	1.E-06	0.6%	3.5%	6.8%	10.4%	14.8%	20.5%
	1.E-07	3.4%	6.3%	9.7%	13.4%	17.9%	23.7%
	1.E-08	5.8%	8.8%	12.3%	16.1%	20.7%	26.6%
	1.E-09	7.9%	11.0%	14.5%	18.4%	23.1%	29.2%
	1.E-10	10.6%	13.8%	17.4%	21.4%	26.2%	32.4%

**Table 2.** The gain due to the activation of an AC mechanism for an activity grade  $\sigma$  of 0.5.

		$P_{session}$					
		1.E-08	1.E-07	1.E-06	1.E-05	1.E-04	1.E-03
$P_{packet}$	1.E-03	-6.2%	-4.3%	-2.2%	0.3%	3.2%	6.7%
	1.E-04	-3.9%	-1.9%	0.3%	2.8%	5.8%	9.4%
	1.E-05	-1.7%	0.3%	2.5%	5.1%	8.2%	11.9%
	1.E-06	0.3%	2.3%	4.6%	7.2%	10.4%	14.1%
	1.E-07	2.1%	4.1%	6.5%	9.1%	12.3%	16.1%
	1.E-08	3.9%	6.0%	8.4%	11.0%	14.3%	18.2%
	1.E-09	5.5%	7.6%	10.0%	12.7%	16.1%	20.0%
	1.E-10	7.1%	9.2%	11.7%	14.5%	17.8%	21.8%

**Table 3.** The gain due to the activation of a CAC mechanism for an activity grade  $\sigma$  of 0.8.

$\sigma = 0.8$		$P_{session}$					
		1.E-08	1.E-07	1.E-06	1.E-05	1.E-04	1.E-03
$P_{packet}$	1.E-03	-3.6%	-2.4%	-1.2%	0.0%	1.6%	3.2%
	1.E-04	-2.4%	-1.2%	0.0%	1.2%	2.8%	4.4%
	1.E-05	-1.2%	0.0%	1.2%	2.4%	4.1%	5.7%
	1.E-06	-0.4%	0.8%	2.1%	3.3%	4.9%	6.6%
	1.E-07	0.4%	1.7%	2.9%	4.1%	5.8%	7.5%
	1.E-08	1.3%	2.5%	3.8%	5.0%	6.7%	8.4%
	1.E-09	2.1%	3.4%	4.6%	5.9%	7.6%	9.3%
	1.E-10	3.0%	4.3%	5.5%	6.8%	8.5%	10.2%

## 6 Conclusion

The results of this study show that for a broad value range of the tolerated session blocking and packet loss probabilities, the application of CAC enables the service provider to serve considerably more subscribers. For a subscriber activity grade  $\sigma$  equal to 0.2, the difference amounts to up to 30%. This advantage of CAC decreases, however, with increasing subscriber activity. Due to a worst-case consideration concerning CAC, and the fact that a model of the packet loss burstiness in the case without CAC is not available, the quantitative comparison could be even more advantageous for CAC than our results indicate. Moreover, in case of exceptionally high service demands, CAC guarantees the quality and continuity of the service by refusing the excess of service requests, whereas in a scenario without CAC, the drastically increased packet losses may cause a complete break-down of the service.

## References

1. Bruneel, H., Kim, B.G., "Discrete-time model for communication systems including ATM" (Kluwer Academic Publishers, Boston, 1993).
2. Cohen, J.W., "The generalised Engset formulae", Philips Telecommunication Review, 1957, vol. 18(4), pp. 158-170.
3. De Neve, P., et al.: A comprehensive quality performance study of MPEG streaming video over an IP-based network with packet loss. Submitted to ICC 2004, 20-24 June, 2004.
4. Bruneel, H., Steyaert, B., Desmet, E., Petit, G.H., "Analytic derivation of tail probabilities for queue lengths and waiting times in ATM multiserver queues", European Journal of Operational Research, 1994, vol. 76, pp. 563-572.
5. Kim, H., Shroff, N.B., "Loss Probability Calculations in a Finite Buffer Multiplexer", IEEE/ACM Trans. on Networking, vol. 9, no. 6, Dec. 2001, pp. 765-768.
6. Mao, G., Habibi, D., "Loss performance analysis for heterogeneous ON-OFF sources with application to connection admission control", IEEE/ACM Transactions on Networking, vol. 10 (1) (February 2002), pp. 125 - 138.
7. Steyaert, B., Bruneel, H., Analytic derivation of the cell loss probability in finite multiserver buffers from infinite-buffer results, Proceedings 2nd IFIP WG6.4 Workshop on Performance Modelling and Evaluation of ATM Networks (1994), pp. 18/1-18/11.
8. Steyaert, B., De Vleeschauwer, D., Queueing models for voice in packet-based networks, Presentation at the COST 257 Final Seminar "Impacts of new services on the architecture and performance of broadband networks" (Würzburg, 27-29 September 2000).