

Performability Analysis of an Adaptive-Rate Video-Streaming Service in End-to-End QoS Scenarios*

I.V. Martín F., J. Alins, Mónica Aguilar-Igartua, Jorge Mata-Díaz

Telematics Engineering Department, Technical University of Catalonia (UPC), Jordi Girona 1-3,
08034, Campus Nord, Barcelona, Spain
{isabel, juanj, maguilar, jmata}@entel.upc.es.

Abstract. Nowadays, dynamic service management frameworks are proposed to ensure end-to-end QoS. To achieve this goal, it is necessary to manage Service Level Agreements (SLAs), which specify quality parameters of the services operation such as availability and performance. This work is focused on the evaluation of Video-on-Demand (VoD) services in end-to-end QoS scenarios. Based on a straightforward Markov Chain, Markov-Reward Chain (MRC) models are developed in order to obtain various QoS measures of an adaptive VoD service. The MRC model has a clear understanding with the design and operation of the VoD system. In this way, new design options can be proposed and be easily evaluated. To compute performability measures of the MRC model, the randomization method is employed. Predicted model results fit well to the ones taken from a real video-streaming testbed.

1 Introduction

During the last years, Video-on-Demand (VoD) applications for the transmission and distribution of video have experienced a growing development and acceptance from the users. Video-streaming systems have a special relevance in wired and wireless networks. In these systems, the video is distributed for its reproduction in real-time [1]. The video server of a video-streaming system stores a set of movies that can be requested by any client. If the connection request is accepted, a session is initiated; then a multimedia stream flows through a set of heterogeneous networks from the video server to the client terminal.

In end-to-end Quality of Service (QoS) scenarios, QoS measures such as packet loss, packet delay and jitter must be guaranteed when the connection is accepted. These real-time guarantees required by the VoD systems could be achieved using QoS differentiation between traffic classes over heterogeneous networks. On the other hand, with the aim of reducing the huge amount of information generated by the video source, loss compression techniques are applied. The most common coding techniques are H.26x and MPEG standards [2]. The price to pay for a high compression level is a degradation level in the image quality.

VoD systems with QoS guarantees might be designed to provide a uniform level of image quality to their users. The video flows coded with a constant image-quality

* This work has been financed by the Spanish investigation projects DISQET (CICYT TIC2002-00818), CREDO (CICYT TIC2002-00249) and ARPA (CICYT TIC2003-08184).

present a high variability in their transmission bit rate. In this way, the amount of network resources required for the transmission fluctuates notably.

The adaptive VoD services employ a set of policies for dynamic resource allocation. It is accomplished by means of signalling protocols used between the service and the network. Thus, related to the bit rate variability of the flow, the service raises renegotiations to the network in order to modify the allocated resources during the session. These renegotiations are performed at the temporal-scale of the scenes in a video sequence. In this way, the amount of network resources reserved during the session are reduced substantially, and a more efficient exploitation of these resources is achieved [1,3,4]. Therefore, the number of concurrent streaming sessions in the system is incremented. However, the image quality will be reduced in some congestion moments when the flow with the selected quality cannot be transmitted. In these congestion situations, the service adapts the transmission bit rate to the available network resources applying a higher compression level or managing the enhanced layers when scalability techniques are employed [1]. Thus, the final QoS provided to the customers of these streaming services depends on the available network resources.

Both the service providers and the customers are indeed interested in tools which support to quantify the performance of these systems from their points of view. Analytical tools are the most appropriate mechanisms to facilitate the required evaluation. Moreover, these tools should provide feasibility to incorporate modifications into the system in an easy way. Further, they also must admit a computational evaluation. This kind of analytic tools help to address some of the typically required main objectives: to maximize the use of network resources and the QoS offered to the users and, to define billing metrics. Likewise, these tools could compute diverse parameters in order to specify, to manage and to control the fulfilment of the Service Level Agreements (SLAs). The management of SLAs is a current challenge into the multimedia services area. Further, it has a great commercial interest. There are diverse recent proposals about SLA management (e.g. [5,6]), although none of them specifies how to quantitatively evaluate the user-level SLA parameters.

One of the main objectives addressed in the present work is to compute *a priori* the QoS offered to the user of a video-streaming application. In particular, we are interested in the evaluation of adaptive VoD systems, in which video sources are capable to adapt their output bit rate to the time-varying available network resources.

Some proposals of design and evaluation of adaptive VoD systems are presented in [3,7,8,9,10]. However, most of these proposals use either simulation models or real platforms to carry out the performance evaluation of these systems. These evaluation techniques hinder the system analysis and also the study of several design options. In addition, some analytical proposals do not regard the interaction between the different video sources sharing the network resources. On the other hand, works focused on characterizing and modelling a single video flow [4,11,12] are not enough to evaluate an adaptive VoD session because they do not take into consideration the dynamism of the video quality changes all over the transmission of the stream.

In [13] we proposed a generic method to develop Performability models [14] for VoD systems. This method solves the lacks above mentioned. The applicability of this method is based on the characterization of the coded multimedia flows and the channel

behaviour. This characterization requires suitable markovian models of both the flows and the channel. In addition, an analytical model developed with this generic method for a VoD service was presented in [13]. This model provides accurate results for the measures of user-level QoS parameters such as the image quality, reserved resources, or effectively-used resources. However, a problem of this model is that the computational cost may increase dramatically when the amount of accepted connections or the amount of user-classes increase.

In the present work, we obtain other two new analytical models based on the method presented in [13]. These models reduce the states space to characterize the resources reserved by a group of users. Using these proposed models, we have analyzed the performance of a VoD service varying some design parameters.

The rest of the work is organized as follows. Section 2 describes the evaluated VoD system. In section 3, a background of the previous work presented in [13] is summarized. In section 4 we propose two new simpler analytical models of the adaptive VoD service. Next, in section 5 some numerical results evaluating both VoD models are shown. The results of these models are compared with experimental measurements obtained from the SSADE project (<http://ssade.upc.es>) implemented by the Telematics Services Research Group of the Polytechnic University of Catalonia. Finally, conclusions and future work are presented in section 6.

2 The System Description

Figure 1 depicts the VoD system analysed in this work. Video sequences have previously been coded using the VBR MPEG-II algorithm and then, stored in the video server. When any customer of the VoD service demands one of these sequences, a connection is established if the video service has enough resources to provide the contracted user's profile, i.e. the agreements specified in the service contract. In many IP QoS-aware networks, the RSVP (Resource reSerVation Protocol) is employed as signalling protocol to manage resource reservation requests [15]. Video-streaming services that use RSVP send requests to the network in order to adjust the required resources of the video-stream transmitted. The description of these resources is specified by means of the *Traffic Specification* (TSpec) parameters carried in the PATH messages of the RSVP. The VoD server requests resource reservation for each session dynamically, i.e. the amount of required resources are calculated and adapted for smaller intervals than the length of the sequence. So, different network resources are requested over the whole transmission of a sequence. Moreover, this renegotiation process yields that the available network resources change due to the interaction between the multiplexed connections. With the aim that the video sources are capable to adapt their output bit rate to the time-varying network resources, each available sequence has a set of MPEG flows coded with different quantization step (Q). Then, each available flow offers a different image quality according to Q [2]. For each accepted session, the transmitted stream will match with one of the different available coded flows of the requested sequence. This selection changes depending on the image quality contracted by the user and on the result of the reservation request produced by the end-to-end admission control of the RSVP-based system.

To carry out the system functions, three blocks have been designed as it is sketched in Fig. 1. These blocks are performed as follows. For each available flow, the *Statistical Planner* block has previously calculated and stored the TSpec parameters of each scene and the events of resources renegotiation in each sequence. When scene changes or variations of the available resources happen, the *Regulator/Negotiator* block decides which flow (Q_i) will be transmitted. To guarantee a minimum video quality, the minimum reservation needed to transmit the lower image quality flow must always be assured. The *Traffic Shaper* block extracts the variability introduced by the frame coding modes (Intra, Predicted and Bidirectional-Predicted) of the MPEG algorithm. In this way, the bit rate is smoothed and it is maintained constant (r_{GoP}) for a GoP (Group of Pictures) interval.

As an example of the transmitted stream, Fig. 2 shows the bit rate required to transmit the sequence "The Graduate" coded with a quantification step Q equals to 4, 8 and 16. The dark line shows the bit rate reserved to the connection, when the total resources for the video service are 270,000 Bytes per second. Notice that this stream matches with one of the available flows for each moment of transmission time.

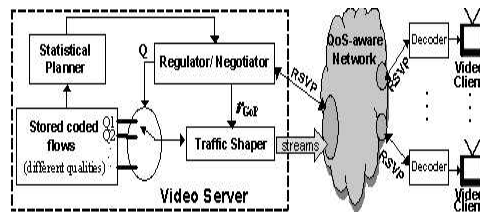


Fig. 1. System model for the VoD service.

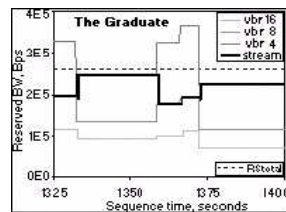


Fig. 2. Required bit rate.

3 Background

3.1 Scene-based Markovian models for a video sequence

In order to efficiently characterize the network resources required by a constant-quality flow of a video sequence, we need to identify the groups of frames with same complexity or activity in the sequence. The identification process of these consecutive groups of frames has been called in the literature as *Segmentation* [16]. The segmentation of a video sequence results on series of groups of pictures with similar requirements of network resources [17]. These segments, also named scenes, define different complexity levels within the sequence. Through the classification of the scenes into activity levels, scene-based models have been proposed in previous works [18]. Some of the more relevant works have developed analytical models based on Markov chains. These models set the number of scene classes heuristically. Straightforward scene-based Markovian models represent scene changes by means of transitions between states, where states identify classes of scenes. For the sake of the simplicity, we will refer to each class of

scenes as an activity level. An example of the Markovian scene-based model is shown in Fig. 3, where L activity levels are defined.

The segmentation process of different constant-quality flows of the same movie gives rise to the same scene bounds. Consequently, for a set of video-flow models related to the same movie, the changes of scene occur at the same time. Fig. 2 remarks how these flows are time-tracked. Hereafter, we will indistinctly refer to a change of the activity level of the sequence as well as a change of the activity level in anyone of its coded flows. In this work, the amount of resources required to transmit the scenes of each available flow are considered to be known and they have previously been calculated for all the stored sequences at the video server.

3.2 Generic Method to Develop Analytical Models of VoD Systems

A generic method to construct Markov Reward Chain (MRC) models for VoD systems was proposed in [13]. Finally, to carry out computations of QoS measures, the method of Randomization [19] was applied to the MRC. The generic modelling methodology presented there consists on 5 steps that obtain a MRC that statistically characterizes the network resources required by a connection and the available resources for this connection. Also, applying this methodology, an analytical model of a particular VoD system was derived in [13]. For more information of this model and the generic method, please refer to [13]. Below, we summarize this model and later on in section 4 we propose some modifications to this model obtaining new models which reduce the space of states and provide a faster computation of expected results.

In **Step 1**, a markovian model for each available fixed-quality video flow in the VoD system is derived in a similar way as it was developed in [3]. This model is composed by three states that define three activity levels (as it is presented in section 3.1).

The Continuous-Time Markov Chain (CTMC) shown in Fig. 4 is an example derived from **Step 2**. For the sake of the clarity, only 3 different constant image-quality flows have been depicted in the draw. Each column corresponds with the model of each available flow, which was obtained in step 1. This CTMC models the behaviour of a connection in the system, where e_f^a is a connection state where flow of quality f (1: worst quality... F : best quality) in the activity level a (0: regular, 1: medium, 2: high) is transmitted through the connection. The transitions between states reflect the scene changes and renegotiation decisions that have been designed in the VoD system. In addition, in this example while the stream remains in the same activity level, the system periodically tries to improve sending requests for the next better image-quality. We call this process *polling of improvement*. The transition rates in this CTMC depend on the required resources and on the available resources. In order to formally express these dependencies, two boolean factors are defined:

$$\alpha(e_f^a) = \begin{cases} 1, & \text{if } RSV(e_f^a) \leq RS_{available} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

and

$$\beta(e_f^a, e_g^b) = \begin{cases} 1, & \text{if } RSV(e_f^a) \leq RSV(e_g^b) \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where $RS_{available}$ is the amount of available resources for the connection and $RSV(e_f^a)$ is the amount of network resources that the system reserves for the connection when the server transmits a flow of quality f in activity level a .

Let $\Psi(e_f^a, e_g^b)$ be the transition rate from state e_f^a to state e_g^b for a connection. These rates in the CTMC, related to increments or decrements of the activity level in the video sequence, are formulated by means of equations which depend on the rates $\lambda_{a,b}$ and the factors β (see [13]). Finally, transitions owing to the *polling of improvement* are formulated as a function that depends on the poll rate λ_p and the factor α . Note that the factors α and β provide the general characterization of the adaptive VoD service by means of the proposed CTMC. Different behaviours for the renegotiation mechanism can simply be designed using these factors. When several connections are being served, $RS_{available}$ vary during the time. In this case, it is mandatory to know the state of all the connections that interact with the one under analysis.

In **Step 3** the model of N accepted sessions is achieved. Firstly, the state of all the connections with a same QoS profile is defined. In this case, that we called homogeneous case, each connection is characterized with the same parameter values of the generic connection model that has been described in step 2. Let $S^* = \{(n_1^0, n_1^1, n_1^2), (n_2^0, n_2^1, n_2^2), \dots, (n_F^0, n_F^1, n_F^2)\}$ be the system state, where each component n_f^a is the number of connections transmitting a flow of quality f in activity level a . Let $S_{+(g,b)}^{-(f,a)}$ be a system state with one more connection transmitting a flow of quality g in activity level b and one less connection transmitting a flow of quality f in activity level a , regarding state S^* . If all transmissions are uncorrelated, only transitions from state S^* to state $S_{+(g,b)}^{-(f,a)}$ can occur. Then, the transition rates $\Psi(S^*, S_{+(g,b)}^{-(f,a)})$ are expressed as a function of rate ψ multiplied by n_f^a . Where, generalizing (1) for N connections and taking into account a conservative admission control, the factor $\alpha(S^*)$ is expressed as follows:

$$\alpha(S^*) = \begin{cases} 1, & \text{if } \sum_{\forall f} \sum_{\forall a} n_f^a \cdot RSV(e_f^a) \leq RS_{total} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

To develop the case of heterogeneous customers, the simpler way is to define the system state as the joining state of each user class, this is: $S = S^1, S^2, \dots, S^C$.

Step 4 is applied to evaluate the performance of a session in the system with other N accepted sessions. Let e_f^a, S be a system state, where e_f^a describes the state of the connection under evaluation and S characterizes $RS_{available}$. The state S may be described e.g. as the one obtained in step 3 or any markovian model of channel capacity.

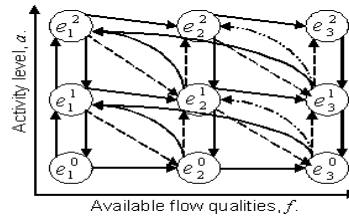
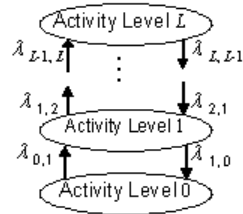


Fig. 3. Scene-based Markovian model. **Fig. 4.** Generic model of an accepted connection.

4 The New Analytical Models

The analytical model we presented in [13] provides accuracy results for the performance measures of the VoD system. Moreover, with this model we accomplish the performance computations and system modifications in an easy way. Nevertheless, as N grows, this model increases combinatorially the number of states and therefore the computational cost increases in the same way [20]. Several methods have been proposed to improve the evaluation efficiency for models with these characteristics [19]. However, these methods may be insufficient to reach the objective of developing a tool to compare different design options of the system at a low calculation time (real time, if possible). To address this issue, we have investigated how to reduce the states space of the model to evaluate the QoS provided to a customer of the VoD system.

In our VoD service, a characterization of the resources reserved by the other N sessions accepted in the system is necessary to provide the available resources for the session under evaluation, $RS_{available}$. This characterization can be reached with the analytical model of N accepted sessions depicted in step 3 of section 3.2, where the states space equals to $(N + 3F - 1)! / (N!(3F - 1)!)$ for F available video-qualities and each ones modelled with 3 activity levels. Further, after applying the step 4 to evaluate the session under analysis, the states space of the model is equal to $3F \cdot (N + 3F - 1)! / (N!(3F - 1)!)$. We have reduced to two states the model of each available flow (see Fig. 3) for the N sessions that are not under analysis. Applying this reduction of states, the states space in step 4 is reduced to $3F \cdot (N + 2F - 1)! / (N!(2F - 1)!)$ for the two models proposed following (see Table 2).

4.1 Analytical Model 1

We have observed that, for any video sequence, the probability to achieve the highest activity level is very low for the model with three activity levels. Based on this observation, we propose to maintain the same regular activity level (level 0) and to establish just one state to define jointly the medium (level 1) and high (level 2) activity levels to model the resources required by each available flow of the sessions that are not under analysis. We do not need to carry out any new statistical analysis of the sequences to develop this new model. We only need to adjust the parameters of the new level 1. We called to this new level as level 1*. The transition rates between level 0 and level 1* are the same ones that between level 0 and level 1 in the analytical model with three levels. The resources required to be in level 1* are calculated as

$$RSV(e_f^{1*}) = \frac{(RSV(e_f^1) \cdot time_medium + RSV(e_f^2) \cdot time_high)}{(time_medium + time_high)} \quad (4)$$

where $time_medium = 1/(\lambda_{1,0} + \lambda_{1,2})$ and $time_high = 1/\lambda_{2,1}$

To characterize the resources reserved by the sessions that are not under analysis, we use the generic method where the model of each available flow includes the modifications above described. This way, we can characterize $RS_{available}$ included in the model of the session under analysis. This latter model is based on three activity levels to characterize each available flow.

We also have evaluated different options beside (4) to approximate $RSV(e_f^{1*})$. For example, $RSV(e_f^{1*})$ equals to $(RSV(e_f^1) + RSV(e_f^2))/2$ and equals to $RSV(e_f^1)$. But (4) has been the option that gives the most accuracy results.

We have compared numerical results of this analytical model with the ones of the original model presented in [13], showing a considerable improvement in the run-time (see Table 2) and the same accuracy. From the analysis of these numerical results, we have been able to adjust some parameters associated to each state leading to the following analytical model.

4.2 Analytical Model 2

We have observed that the regular level is the level with the longest scene length. Therefore, this level is the most influential in the evaluation measures. In addition, the standard deviation of the required resources of the regular level is higher than for the other levels. Then, we propose to modify the parameter $RSV(e_f^0)$ of the model as follows. For each flow in the regular activity level (level 0), a threshold is established to reach a new classification of the scenes in this activity level. Depending on the activity level of these scenes, we have the set of scenes which required resources are lower than the threshold and the set of scenes which required resources are higher than the threshold. We called these new sets of scenes as the lower and the higher sublevel of the regular level, respectively. The amount of required resources associated to the regular level may be any statistical measure of the required resources of all the scenes classified into one of the both sublevels: the lower or the higher. The selection of suitable statistical measures will lead to an analytical model that offers an lower bound or a upper bound for the evaluation measures of our VoD system. We call level 0* to this new regular level.

From the statistical study of different sequences, we have empirically established that best computation for the threshold is to average the required resources for all the scenes classified into level 0, weighted by their scene lengths. In addition, we computed this statistical measure in each resulting sublevel. With this choice, obtained results, compared to the ones from the real system, are very adjusted. However, we are taking additional mathematical methodologies into consideration for future works.

5 Numerical Results

In this section, numerical examples from the evaluation of the VoD system are presented. Several measures of the QoS offered to a customer of the VoD service have been computed, such as the PSNR, the failure time, the transmitted bit-rate and the reserved BW. The expected mean value of these measures for an observed user and the standard deviation between all the users are presented as well. The computation of these measures from the analytical models has been reached using the solution of the moments of cumulative reward presented in [20]. These results have been compared with experimental values from the SSADE video distribution testbed that provides an adaptive video-on-demand service. The configuration parameters of the VoD system used here are the same ones that in [13]. Likewise, the transmitted sequence is “The Graduate”, where the available flows are coded with quantization step Q equals to 4,

8 and 16. For simplicity, we refer to the model presented in [13] as **model_old** which parameters were defined in [13]. We use **model1** and **model2** to refer the two new analytical models presented in sections 4.1 and 4.2, respectively. Table 1 summarizes the values of the modified parameters used for **model1** and **model2**, this is level 0* and level 1* for each available flow of the sequence “The Graduate”. The lower sublevel of the regular level has been chosen as level 0*.

Figures 5 to 7 illustrate some evaluation results of the proposed analytical models. The measurements taken from our testbed are pointed out as **SSADE**. Figure 5 depicts the *Mean PSNR* provided to a user and the standard deviation of this measure for all the accepted sessions. These figures are represented as a function of the total BW assigned to the VoD service and for a variable number of accepted streaming sessions (**N**). For each **N**, the curves start at the minimum BW required to accept the **N**th session in the VoD system. Discontinuities in the analytic curves are produced as a result of the discretization of the activity levels of the flows. Therefore, the performability results are discretized and softened as **N** grows. This happens since the connections are multiplexed and the values associated to the MRC states are softened as **N** increases. The light grey values (**SSADE**) depict measurements of each one of all the sessions accepted in our testbed. Note that, analytical curves of both the **model1** and the **model2** provide a good approximation as a lower and a higher bound with respect to the values from **SSADE**. Both the reserved and the transmitted rates present a similar behaviour (for space reasons they are not shown here).

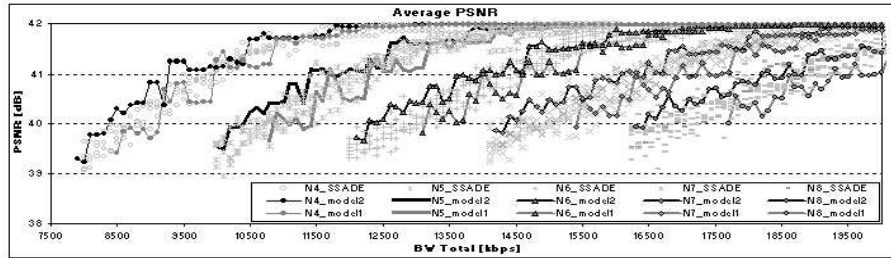


Fig. 5. Mean PSNR provided to a customer for the transmission of “The Graduate” movie

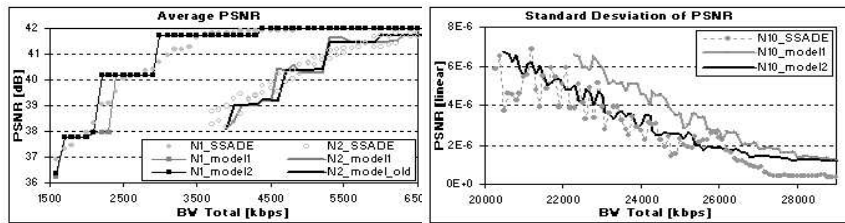


Fig. 6. Mean PSNR provided to a customer.

Fig. 7. PSNR standard deviation.

Table 1. Estimated required resources¹ [bits/GoP]

	level 0*	level 1*
Vbr4	504715.88	782343.95
Vbr8	364476.06	421790.50
Vbr16	364476.06	364476.06

Table 2. Comparison of run-times² [seconds] and states space for **N** sessions, $F=3$, in the system

	model_old		model1 and model2	
N	#states	run-time	# states	run-time
5	4455	~30	1134	~5
8	218790	~950	11583	~60
12	1133730	~35000	55692	~740

The numerical results from the models are also sufficiently accurate to analyze the behaviour of the VoD service from the point of view of standard deviation. Fig. 7 presents the standard deviation for PSNR measurement of the 10 customers in the VoD service. Generally, the evaluation results from **model1** are very close to the ones from **model_old** for any value of **N** (e.g. see Fig. 6). However, run-time of 1st and 2nd moments for **model1** is considerable lower than run-time for **model_old**, as shown in Table 2, without losing precision. Another QoS measure that interests to both customers and service providers, is the total time when the service does not provides the committed QoS level to the customer. This measure can be interpreted as *failure time* (e.g. the service is in failure when the delivered flow is coded with a quality lesser than the video-quality contracted by the customer). Fig. 8 shows numerical results of this measure for a session for 5 and 11 accepted sessions, where a failure is defined when the service does not deliver the flow coded with $Q=4$ (i.e., the maximum video-quality available by the users of our VoD service) to the session under evaluation. In Fig. 8 the total failure time is normalized to the total time of the transmitted sequence. The **SSADE** curve depicts this measure for the last session accepted in the system. For a given total BW assigned to the video service, evidently the *failure time* increases as **N** increases. On the other hand, when the total BW assigned to the video service is the minimum such as **N** sessions can be accepted, we observe that the *failure time* decreases if **N** increases. For example; when $N=5$, total BW=10Mbps, the service is in failure approximately 40% of the total transmission time (see left graph in Fig.8). When $N=11$, total BW= 21Mbps, this time is 30% (see right graph in Fig. 8). Notice that the QoS provided by this VoD system never experiments a high degradation level. The designed user profile is such that each accepted connection can access to all the available resources, without distinction among the users. So the connections quickly acquire a high level of resources sufficient to reach the maximum quality. Therefore, the admission control designed for this VoD service provides suitable QoS guarantees and a strictly access to the system. In this way, once the connection is accepted, the user will perceive a video quality close to the maximum.

Other VoD system configurations or heterogeneous user profile yield other service behaviours. For example, Fig. 9 presents the *Mean PSNR* and reserved BW provided to a session in the system described above (pointed out as **system1** in Fig. 9) where three coded flows are available ($Q=4, 8$ and 16) and an other system which difference is that the flow coded with $Q=8$ is not available (pointed out as **system2** in Fig. 9). The curves in Fig. 9 verify that *mean PSNR* is lower in **system2** than in **system1** for the same reserved BW in both systems. We can establish that the VoD **system1** has a higher efficiency than the VoD **system2**, since it provides a more adaptive service.

¹ The level 1* is used by **model1** and **model2** and the level 0* is only used by **model2**.

² The solutions are implemented on Delphi 4 and executed on Pentium 1.4GHz, 768MB.

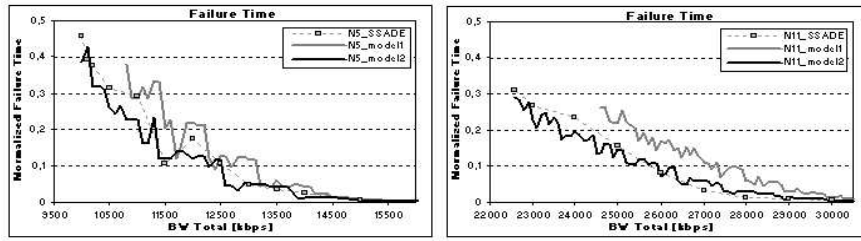


Fig. 8. Failure Total Time for 5 and 11 accepted sessions.

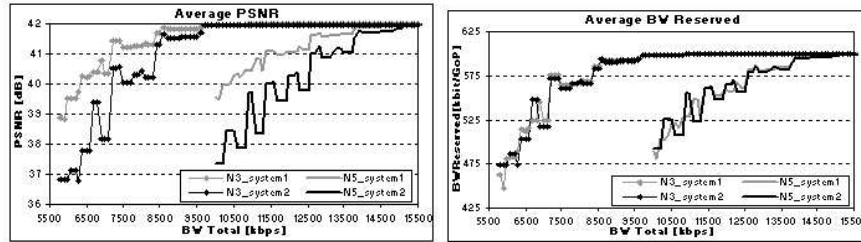


Fig. 9. Comparison between different VoD services (see text for explanation).

6 Conclusions and Future Work

The end-to-end QoS provisioning in IP-based networks is a current challenge. Many QoS management frameworks available in the literature need additional mechanisms and procedures to quantitatively evaluate the user-level SLA parameters. The generic methodology proposed in [13] utilizes the characterization of the available constant-quality flows with the aim to develop an adaptive model for the video-streaming service. Models obtained with this methodology capture the dynamism of the transmitted stream taking into account the activity variations of the sequences and the interaction with the network. In the present work we propose two new analytical models of an adaptive VoD system employing this methodology. These models are simpler than the example presented in [13]. The computational cost is considerably reduced for both new models. The key of the goodness of these models has been to reduce the states space to characterize the resources reserved by a group of users. Different measures of performance evaluation of the VoD service have been calculated using these models. The precision of computed results allow us to verify that the developed models give a good estimation of the performance of adaptive multimedia systems in end-to-end QoS scenarios. The adjustment of the obtained results will depend on the suitable characterization of the multimedia sources and the transmission channels using suitable Markovian models. As future lines of research, we are studying others efficient ways to develop the heterogeneous problem, where diverse QoS customer profiles are considered. Likewise, well-known models in the literature can be integrated to characterize time-varying available resources produced by variations in wireless channel capacity.

References

1. Wu, D., Hou, Y., Zhu, W., Zhang, Y., Peha, J.: Streaming Video over Internet: Approaches and Directions. *IEEE Trans. On Circuits and Systems for Video Technology* **11** (2001)
2. Ghanbari, M.: Video Coding: An Introduction to Standard Codecs (IEE Telecommunications Series 42). IEE Publishing (1999)
3. De la Cruz, L.J., Mata, J.: Performance of Dynamic Resource Allocation with QoS Guarantees for MPEG VBR Video Traffic Transmission over ATM Networks. In: Proceedings of the IEEE GLOBECOM'99, IEEE Communications Society. (1999)
4. Manzoni, P., Cremonesi, P., Serazzi, G.: Workload Models of VBR Traffic and Their Use in Resource Allocation Policies. *IEEE/ACM Transactions on Networking* **7** (1999)
5. Cortese, G., Cremonese, P., D'Antonio, S., Diaconescu, A., Esposito, M., Fiutem, R., Romano, S.P.: CADENUS: Creation and Deployment of End-User Services in Premium IP Networks. *IEEE Communications Magazine* (2003)
6. IST Project: TAPAS– Trusted and QoS-Aware Provision of Application Services (2001) <http://www.newcastle.research.ec.org/tapas/>.
7. Muntean, G., Murphy, L.: A New Adaptive Multimedia Streaming System for All-IP Multi-Service Networks. *IEEE Transactions on Broadcasting* **50** (2004)
8. Lombardo, A., Schembra, G.: Performance Evaluation of an Adaptive-Rate MPEG Encoder Matching IntServ Traffic Constraints. *IEEE/ACM Transactions on Networking* **11** (2003)
9. Luna, C., Kondi, L., Katsaggelos, A.: Maximizing User Utility in Video Streaming Applications. *IEEE Trans. on Circuits and Systems for Video Technology* **13** (2003)
10. Ramanujan, R.S., Newhouse, J., Kaddoura, M., Ahamad, A., Chartier, E., Thurber, K.: Adaptive streaming of MPEG video over IP networks. In: Proceedings 22nd Annual Conference on Local Computer Networks, IEEE. (1997)
11. Adas, A.: Traffic Models in Broadband Networks. *IEEE Communications Magazine* **35** (1997)
12. De la Cruz, L.J., Fernández, M., Alins, J., Mata, J.: Bidimensional Fluid Model for VBR MPEG Video Traffic. In: 4th International Conference on Broadband Communications, IFIP, TC6/WG6.2. (1998)
13. Martín, I.V., Alins, J., Aguilar-Igartua, M., Mata, J.: Modelling an Adaptive-Rate Video-Streaming Service Using Markov-Rewards Models. In: Proc. of the First International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE04), IEEE. (2004)
14. Meyer, J. Teletraffic Science for Cost-Effective Systems, Network and Services, ITC-12. In: Performability Evaluation of Telecommunication Network. Elsevier Science Publishers B. V. (North Holland) (1989)
15. Bernet, Y.: RFC 2998: A framework for integrated services operation over diffserv networks (2000)
16. Sarkar, U., Ramakrishnan, S., Sarkar, D.: Study of long-duration MPEG-trace segmentation methods for developing frame-size-based traffic models. In: Computer Networks. Volume 44. (2004)
17. Wu, M., Joyce, R.A., Wong, H., Guan, L., Kung, S.: Dynamic Resource Allocation via Video Content and Short-Term Traffic Statistics. *IEEE Transactions on Multimedia* **3** (2001)
18. Mashat, A., Kar, M.: Performance Evaluation of a Scene-based Model for VBR MPEG Traffic. *Performance Evaluation IFIP WG7.3* **36** (1999)
19. Haverkort, B.R., Marie, R., Rubino, G., Trivedi, K.: Performability Modelling. Techniques and Tools. John Wiley & Sons (2001)
20. Vallejos, R., Barria, M.: Evaluation of Moments of Cumulative Reward in Repairable Systems (2005) Submitted to the PERFORMANCE 2005 Conference, Jean le Pins, France.