Performance of a partially shared buffer with correlated arrivals

Dieter Fiems, Bart Steyaert and Herwig Bruneel

SMACS Research Group, Department TELIN
Ghent University, St-Pietersnieuwstraat 41, 9000 Gent, Belgium

Abstract. We assess the performance of a partially shared bottleneck buffer for scalable video. The arrival process of the video packets is modelled by means of a two-class discrete batch Markovian arrival process. Using a matrix-analytic approach, we retrieve various performance measures. We illustrate our approach by means of a numerical example.

1 Introduction

Scalable video coding is able to cope with bandwidth fluctuations in packet networks [1]. A video stream is encoded in a base layer and one or more enhancement layer streams. Only the base layer is needed to decode and playback the video, although at a poor quality. Combined with the enhancement layers, the video can be played back at full quality. Intermediate network nodes should therefore drop packets of the enhancement layer to ensure delivery of base layer packets. In this way, the video quality can be reduced gracefully during congestion periods.

To ensure delivery of base layer packets when the network is congested, network nodes are required to support some sort of Quality of Service differentiation. Partial Buffer Sharing (PBS) implements service differentiation by means of a threshold based packet acceptance policy. As long as the number of the packets in the buffer does not exceed a fixed threshold, both base (class 1) and enhancement layer (class 2) packets are accommodated by the buffer. Once the number of packets in the buffer exceeds the threshold, only base layer packets are accepted. A PBS acceptance policy offers space priority at the cost of some overall throughput loss. However, it is easily implementable in practice as opposed to e.g. a push-out buffer [2].

2 Performance analysis

The bottleneck buffer under consideration operates synchronously, i.e., time is slotted. There are two traffic classes (class 1 and 2) and packets of these classes arrive in accordance with a two class discrete-time batch Markovian arrival process (2-DBMAP). Such a process is completely characterised by a doubly indexed sequence of substochastic matrices $A_{n,m}$. The matrix $A_{n,m}$ governs the transitions of the Markovian environment of the 2-DBMAP when there are n class 1 and m class 2 arrivals. The transmission time of packets is fixed and equal to the slot length. Up to N packets can be transmitted at a slot boundary and the buffer can accommodate up to M packets, including the packets being transmitted. However, a class 2 packet is only admitted when there are no more than a threshold $T \leq M$ packets present upon arrival of this packet in accordance with the PBS acceptance policy as described in the introduction. Since there may be arrivals of both classes as well as departures at a slot boundary, one needs to specify the order in which these take place. We here assume the following order: (1) departures; (2) arrivals of class 1; (3) arrivals of class 2.

We consider the queue content at random slot boundaries and let π_i denote the vector whose nth element equals the probability that the 2-DBMAP is in state n at a random slot boundary while there are i packets in the queue at that boundary. By means of matrix analytic techniques, one can show that the vectors π_i satisfy the following set of equations,

$$\pi_0 = \sum_n \pi_n D_{n,N-n}\,, \quad \pi_i = \sum_n \pi_n C_{n,N+i-n}\,, \quad \sum_n \pi_n\, e = 1\,. \tag{1}$$

for $i = 1, 2, \ldots, M-N$. Here e is a column vector of ones and the matrices $C_{n,m}$ and $D_{n,m}$ are defined as follows,

$$C_{n,m} = \sum_{g,h=0}^{\infty} A_{g,h}\, \Theta(g,h;n,m)\,, \quad D_{n,m} = \sum_{k=0}^{m} C_{n,k}\,, \tag{2}$$

with $\Theta(g,h;n,m) = 1(\min(g, M-n) + \min(h,(T-n-g)^+) = m)$ and with $1(\cdot)$ the standard indicator function. One easily shows that this set is of $M/G/1/N$ type and therefore (1) can be solved efficiently by the reduction algorithm [3].

Given the vectors π_i, we may obtain various performance measures. E.g., the accommodated class 1 and class 2 arrival load are given by,

$$\tilde{\rho}_1 = \sum_{n,m} \pi_n m \sum_{k,l} A_{k,l}\, e\, 1(\min(k, M-n) = m)\,, \tag{3}$$

$$\tilde{\rho}_2 = \sum_{n,m} \pi_n m \sum_{k,l} A_{k,l}\, e\, 1(\min(l,(T-n-k)^+) = m)\,, \tag{4}$$

respectively. The class i packet loss ratio (plr) is then given by $\text{plr}_i = 1 - \tilde{\rho}_i/\rho_i$. Here ρ_i denotes the class i arrival load,

$$\rho_1 = \tau \sum_{n,m} n A_{n,m} e\,, \quad \rho_2 = \tau \sum_{n,m} m A_{n,m} e\,, \tag{5}$$

with τ the normalised Perron-Frobenius eigenvector of the matrix $\sum_{k,l} A_{k,l}$.

The analysis of the probability mass function of the packet delay is more involved. Consider a random slot boundary and let $c(l,n,m)$ denote the probability that there are l packets in the buffer and that there are n class 1 and m class 2 packet arrivals that the buffer can accommodate. We have,

$$c(l,n,m) = \sum_{g,h} \pi_l A_{g,h} e\, 1(\min(g, M-l) = n \wedge \min(h,(T-l-g)^+) = m)\,. \tag{6}$$

Given these probabilities, we find that the fraction of slots $\nu_i(k)$ where there is a class i arrival that finds k packets in the buffer upon arrival equals,

$$\nu_1(k) = \sum_{l=0}^{\min(k,N)} \sum_{n=k-l+1}^{M-l} \sum_m c(l,n,m)\,, \tag{7}$$

$$\nu_2(k) = \sum_{l=0}^{k} \sum_{n=k-l+1}^{(T-l)^+} \sum_m c(m,l-m,n)\,, \tag{8}$$

for $k = 0,1,\ldots,M-1$ and for $k = 0,1,\ldots,T-1$ respectively. As there is at most one such packet arrival at a slot boundary, the probability $u_i(k)$ that a random class i packet finds k packets upon arrival in the buffer equals $\nu_i(k)/\tilde{\rho}_i$. Finally, let packet delay be defined as the number of slots between a packet's arrival and departure slot boundary. Since up to N packets leave the buffer system at a slot boundary, the probability $d_i(n)$ that the delay of a class i packet equals n slots is given by,

$$d_i(n) = \sum_{k=(n-1)N}^{nN-1} u_i(k)\,, \tag{9}$$

for $n = 1,2,\ldots,\lceil M/N \rceil$ and for $n = 1,2,\ldots,\lceil T/N \rceil$ for the class 1 and class 2 delay respectively.

3 Numerical Example

To illustrate our approach, consider the case where 8 video sources are routed through a bottleneck buffer ($M = 100$, $N = 4$). Playout buffers are used at the destination nodes to cope with delay fluctuations in the network. Each video source is modelled as an on/off source generating one packet at each slot boundary in the on state and no packets in the off state. A fraction θ of the packets belong to class 1. The on/off processes are completely characterised by the pair (σ, K). Here σ denotes the fraction of time that the source is on and K is measure for the absolute lengths of the on- and off-periods [4]. In the remainder we set $\sigma = 0.4$, corresponding to an 80% load of the bottleneck buffer.

In Fig. 1(a), plr_1 and plr_2 are depicted vs. T for various values of K and for $\theta = 0.5$. For each K, the upper and lower curves depict plr_1 and plr_2 respectively. For $K = 10$, the middle curve depicts the plr of a random packet. Increasing T yields an exponential decrease of plr_2 at the cost of an exponential increase of plr_1. Also, the plr of a random packet increases for decreasing values of T. I.e., PBS offers service differentiation at the cost of additional packet loss. Further, performance of the buffer system deteriorates when the arrival process is more bursty (larger K). Fig. 1(b) depicts the experienced packet loss ratio of the class 1 (eplr_1) and 2 (eplr_3) video flows versus the playout delay δ for various values of θ. The eplr_i includes packet loss in both bottleneck and playout buffer (due to underflow). I.e., a packet is not lost if it is not dropped by the bottleneck buffer

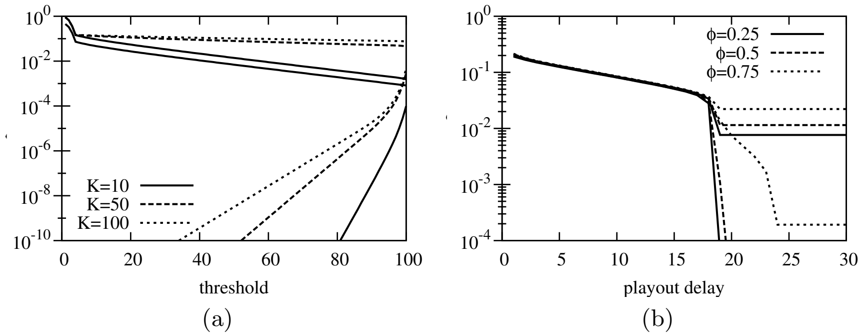

Fig. 1. Class 1 and 2 packet loss ratio vs. threshold T (a) and experienced packet loss ratio of class 1 and 2 vs. playout delay.

and if its delay in this buffer does not exceed δ. The parameters T and K are set to 80 and 100 respectively. For small δ, loss is mostly caused by buffer underflow in the playout buffer. Once δ reaches $\lceil T/N \rceil$, class 2 underflow is no longer possible in the playout buffer since the delay in the bottleneck buffer is bounded by $\lceil T/N \rceil$. We have, $\mathrm{eplr}_2 = \mathrm{plr}_2$. Also eplr_1 drops fast. This is explained by the fast decay of the class 1 delay probability mass function.

Acknowledgement.

This work has been partly carried out in the framework of the CHAMP project sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

References

1. Radha, H., Chen, Y., Parthasarathy, K., Cohen, R.: Scalable internet video using MPEG-4. Signal Processing: Image communication **15**(1-2) (1999) 95–126
2. Kröner, H., Hébuterne, G., Boyer, P., Gravey, A.: Priority management in ATM switching nodes. IEEE Journal on Selected Areas in Communications **9**(3) (1991) 418–427
3. Blondia, C., Casals, O.: Statistical multiplexing of VBR sources: A matrix-analytic approach. Performance Evaluation **16** (1992) 5–20
4. Fiems, D., Steyaert, B., Bruneel, H.: Analysis of a discrete-time GI-G-1 queueing model subjected to bursty interruptions. Computers & Operations Research **30**(1) (2003) 139–153