On the Efficiency of Shaping Live Video Streams

György Dán and Viktoria Fodor
KTH, Royal Institute of Technology
Electrum 229, 16440 Kista, Sweden
{gyuri,viktoria}@it.kth.se

Abstract

In this work the efficiency of shaping live video streams is considered. We propose low complexity shaping algorithms adequate for real-time operation and supporting applications with a wide range of delay tolerance. The effect of shaping is investigated considering video streams multiplexed at an output link with a small buffer to absorb packet scale congestion. The advantage of using small buffers when transmitting video streams is the limited delay and delay variation. Consequently, we concentrate on the loss characteristics to evaluate the performance of the proposed solutions. We present mathematical analysis based on fluid flow modeling and the theory of large deviations and confirm the results with simulation.

Keywords: quality of service, live video transmission, source shaping, packet scale buffering, large deviation theory

1 INTRODUCTION

The transmission of live video traffic over the Internet is a fundamental problem of network design, since many video applications require limited end to end packet loss, delay and delay variation. It is generally accepted that traffic control functions must be employed to guarantee these service requirements at a reasonably high network load. The introduction of new control functions in the Internet, however, is a critical issue. First, a very high number of networking devices has to be updated or replaced, second, the complexity of the control functions may limit the span and the transmission capacity of the network.

Recently research efforts address this question by proposing probe based endpoint admission control (PBAC) solutions [6, 3] to provide quality of service guarantees in a way that the functionalities of the routers are kept simple and traffic control functions are placed into the hosts – or edge gateways – only. In the PBAC schemes a host, before transmitting traffic with QoS requirements, probes the network's transmission capability by sending a sequence of probe packets and decides about the transmission based on the statistical quantities of the probing process.

While the admission control ensures that the load of the network stays reasonably bounded, additional control functions can be applied at the hosts to increase the acceptable load, like traffic shaping to decrease the burstiness of the traffic streams and thus decrease the packet loss at the multiplexing nodes and forward-error correction to recover from packet losses.

In this work we investigate how source shaping can increase the efficiency of live video transmission considering MPEG video streams multiplexed at an output link with small buffer. The advantage of using small buffers is that the delay and delay variation is strictly limited by the buffer size an thus only the packet loss has to be controlled. Source shaping provides the following favorable properties: i) it does not require global decision or any modification in the network, ii) can improve the service quality of streams with different QoS requirements and traffic characteristics and iii) can be introduced in the network gradually.

We propose two solutions to shape video streams without packet loss and with given delay bound. The first scheme follows the ideas presented in [10], and determines the shaper rate considering the delay bounds of all the frames waiting in the shaper buffer. The second scheme considers only the delay bound of the last frame in the buffer, thus provides a solution with very low computational complexity.

We evaluate the performance of the proposed schemes with simulation and with analytical methods based on fluid flow approximation applying the theory of large deviations [1]. As the delay and delay variation is limited by the buffer size, the analysis focuses on the packet loss characteristics, as average loss, the loss of shaped and unshaped streams and the distribution of packet losses among the frame types of the MPEG stream.

The paper is organized as follows. In the next section we discuss related works and results. Section 3 describes the system model with the sources, the shapers and the multiplexer. Section 4 explains the two shaper algorithms we propose, and Section 5 presents the analytical model to evaluate the efficiency of source shaping. In Section 6 we present and discuss numerical results and in Section 7 we conclude our work.

2 PREVIOUS WORK

In this section we survey previous research results that inspired our work on shaping and transmission of on-line video traffic over the Internet.

Traditional solutions to provide QoS guarantees in packet switched networks are based on per flow reservation messages and capacity reservations (e.g., RSVP), and as a consequence, suffer from scalability limitations [17]. To circumvent these scalability problems, several recent works have proposed some form of probe based endpoint admission control (PBAC). In these solutions the hosts (endpoints) send a sequence of probe packets before user data transmission, to detect the level of congestion in the network. The level of congestion and thus the possibility of user data transmission with the required QoS parameters is determined from the statistical quantities of the probe transmission process, like probe loss probability [6], delay and delay variation [2] or packet marking probability [5]. An excellent evaluation of the various designs can be found in [3]. The PBAC schemes provide QoS guarantees without the need of control functions inside the network, support QoS requirements depending on the users needs, and do not require the complex description of the traffic streams. Our work follows the basic idea of these solutions by investigating how additional control functions at the host can increase network efficiency.

The transmission of video streams requires limited packet loss, end to end delay and delay variation. Obviously, these values depend on the size of the buffers at the routers. Depending on the size of the buffer one can differentiate between packet scale buffering and burst scale buffering. In the first case only a small buffer is provided to absorb packets arriving simultaneously, thus the packet loss probability might be high while the delay is strictly limited. In the second case the buffer provides enough space to absorb larger bursts and consequently, limits the loss probability while the control of delay and delay variation becomes a complex issue [14]. The use of packet scale buffering for transmitting delay and loss sensitive data like coded video streams have been proposed in [6, 15, 16], showing that low packet loss probability and high network utilization can be achieved if the peak rate of the streams is low compared to the link capacities. In [15, 16] packet scale buffering is proposed together with source shaping. It is proved that a single buffer leaky bucket is an optimal shaper in this scenario.

The shaping of stored variable bit rate video streams is widely analyzed in the literature. Recent solutions are based on network calculus e.g., [4, 11] or the bounding interval dependent (BIND) characterization of the streams [7, 8]. Only a few works address the shaping of live video streams. The main questions to face in this case are the tradeoff between the delay introduced at the shaper and the available information on the traffic to be transmitted and the effectiveness and the complexity of traffic prediction. A solution for shaping with limited packet loss is proposed in [9]. The shaping is based on the BIND characterization. The BIND parameters are continuously updated as the statistical properties of the stream change, which seems to be a rather complex process considering the real-time operation. In [12] the authors study statistically identical, peak rate controlled and leaky bucket shaped sources feeding a buffered multiplexer. The goal is to find the shaper rate that minimizes network resources like buffer and bandwidth, while keeping the delay limited. The solution, however, can not handle multiplexed sources with different characteristics. Algorithms for lossless shaping of individual streams are presented in [10] and [13]. In these works the statistical quantities of the shaped streams are analyzed, but network scenarios are not considered. In [13] shaping with delays in the range of 2-30 seconds is proposed, an adequate solution for broadcasting applications. The algorithm shown in [10] works with shaper delays less than a second. To avoid the fluctuation of the shaper rate, the algorithm uses past frame sizes to predict future traffic intensity, and needs to know the length of a GOP in advance.

Our contribution to this line of works is the definition of shaping algorithms with very low complexity that support the transmission of live video streams with a wide range of acceptable shaper delay and the analysis of the effect of shaping when the video streams are multiplexed at an output link with small buffer. To the best of our knowledge, such results have not yet been presented in the literature.

3 MODEL DESCRIPTION

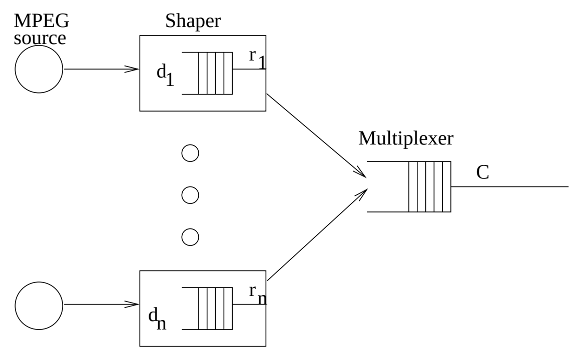

Figure 1: The considered system with MPEG sources, source shapers and a multiplexer.

The system model considered in this paper is shown in figure 1. The system includes traffic sources, source shapers and a multiplexing node with a single output link.

The sources generate MPEG coded streams, the most commonly used encoding scheme for the storage and transmission of video information. MPEG is a family of standards used for coding visual information in a digital compressed format. It has been designed to support a broad range of transmission rates and hence a broad range of visual quality. In an MPEG stream information is stored as a sequence of frames, corresponding to a sequence of pictures in a video, generated with fix time intervals. Compression is achieved by eliminating the spatial and temporal redundancy of the information in the frames. Spatial redundancy is decreased by intraframe coding of the individual frames, while temporal redundancy is reduced by interframe coding between subsequent frames. Thus the sequence of frames consists of intraframe coded frames (I frames), and interframe coded predicted (P frames) and bidirectionally predicted frames (B frames). The subsequent frames between two consecutive I frames form a group of picture (GOP). The GOP structure of the streams can be different, depending on the required quality. A typical example for the sequence of frames is IBBPBBPBBPBB. As a consequence of the coding scheme, information loss in the three frame types has different effect on the percepted visual quality. The loss of data in an I frame propagates forward through the next GOP and backward to the last P frame (affecting up to 14 frames if the number of frames in an open GOP is 12). Meanwhile, the loss of data in a B frame only affects that particular frame.

The intra- and iterframe coding results in the fluctuation of the frame sizes on two timescales. The intraframe coding compresses complex scenes with less efficiency, and consequently, the frame sizes change on the long term at the scene changes. The interframe coding leads to short term frame size fluctuation, since I frames are usually significantly larger than P frames, and P frames are larger than B frames. The scale of the fluctuation is about a factor of 3 on the long, and a factor of 10 on the short term.

Shapers are used at the sources to decrease the frame to frame fluctuation of the coded video stream. The shaper we use in this work is a single buffer leaky bucket, as it is proved to be optimal for networks with small buffers [15]. Frames leaving the encoder are stored in the shaper buffer and are transmitted with a given transmission rate. The shaper is designed to introduce limited delay and provide lossless transmission, that is, no data can be lost due to buffer overflow or delay limit violation. To achieve this, the shaper transmission rate has to be adjusted depending on the size of the arriving frames and the buffer size has to be large enough to store the frames waiting for transmission. If the frames arrive at a regular basis, the maximum number of frames in the buffer can be bounded by the ratio of the delay limit to the frame interarrival time.

The shaped video streams are multiplexed at a network node with a single output link. Since the arrival rate of the streams can temporarily exceed the capacity of the outgoing link, the node is equipped with a buffer to store arriving data. In this work we consider packet scale buffering, the size of the buffer is in the order of the ratio of the output link transmission capacity to the peak rate of the video streams.

4 SOURCE SHAPING ALGORITHMS FOR LIVE VIDEO STREAMS

In this section two algorithms are proposed to control the rate r of the source shaper when transmitting live MPEG video streams. Both aim to minimize the maximum and the variance of the transmission rate, and fulfill the following requirements:

i) The transmission delay in the shaper does not exceed the predefined maximum shaper delay and the shaping is lossless.

ii) The algorithms are simple in terms of the complexity of the shaper rate calculation and the amount of information considered, a requirement to assist real-time operation.

iii) The algorithms provide efficient solution for shaping with a large range of shaper delays.

iv) The algorithms do not require any apriori information on the encoding scheme of the video, i.e., the number and sequence of I, P and B frames in a GOP.

Both of the algorithms assume that the shaper can detect the type of the arriving frame. In general, the following rules apply to select the shaper rate r:

1. The shaper rate can be changed at any time t when a new frame is generated and placed into the shaper buffer.

2. The shaper rate is increased if the new frame can not be transmitted within the delay limit d.

3. The shaper rate is decreased if the new frame is of type I, and all the traffic in the shaper buffer can still be transmitted within the delay limit. This rule is based on the assumption that small P or B frames do not indicate intensity change in the video stream. To decrease fluctuation, the new shaper rate is calculated as the average of the current rate and the minimum rate allowed by the delay limit.

4. P and B frames entering the shaper when the buffer is empty are transmitted with a rate such that the frame leaves the buffer before the new frame arrives, i.e., in one frame time, in order to prevent the shaper from keeping data before larger I and P frames arrive.

The two proposed algorithms described in the following differ in applying rule 3. The first algorithm is optimal in the sense, that the residual acceptable delay is considered for all the frames stored in the shaper buffer to determine the minimum shaper rate. The second, simplified algorithm does not follow the delays of the individual frames in the buffer, and calculates the shaper rate based on the buffer content only.

Shaper rate control based on residual transmission delays

The algorithm based on residual transmission delays (RTD) works as follows. For every frame entering the shaper, the size of the frame f_i and its latest departure time $t_i = t + d$ is recorded. The minimum shaper rate $r_{min}(t)$, allowed by the delay limit is calculated as

$$r_{min}(t) = \max_{n<N} \frac{f_0(t) + \sum_{j=1}^{n-1} f_{i-N+j+1}}{t_{i-N+j+1} - t}, \qquad (1)$$

where N is the number of frames in the shaper at time t and $f_0(t)$ is the residual size of the first frame in the buffer at time t. The residual size is less than the original frame size if the transmission of the frame has already started.

The complexity of the shaper rate calculation is $O(N^2)$ additions and $O(N)$ divisions, where the value of N is bounded by d/T_{frame}. In addition to the actual shaper rate, the shaper has to remember the size and the arrival time of the frames waiting for transmission. A system clock has to be maintained and read at each frame arrival.

Shaper rate control based on the buffer content

This solution does not record the residual acceptable transmission delay for the frames waiting in the shaper buffer, the shaper rate calculation is based on the buffer content (BC).

When frame i arrives to the shaper, its size is added to the amount of data in the shaper $b(t) = b'(t) + f_i(t)$, denoting the buffer occupancy before the frame arrival as $b'(t)$. The minimum shaper rate is calculated considering the buffer occupancy at the time of the new frame arrival:

$$r_{min}(t) = \frac{b(t)}{d}, \tag{2}$$

To avoid delay bound violation for frames stored in the buffer, the shaper rate can be decreased only if the buffer is empty before the new frame arrival (i.e., b'(t)=0), a significant constraint on rule 3 above.

As a consequence, this simplified BC algorithm follows the decreasing intensity of the stream with some delay compared to the RTD solution. The complexity of the BC algorithm, however, is very low (one addition and one division at each frame arrival), there is no need for system clock information and only the number of bytes waiting in the shaper buffer has to be stored.

5 ANALYTICAL MODEL

In this section we present an analytical method to calculate the overall packet loss probability and the distribution of the packet losses among sources at a multiplexer performing packet scale buffering.

The analysis is based on the fluid flow modeling of the traffic streams and uses results of the theory of large deviations to approximate probabilities of rare events. A short summary of the basic ideas behind the large deviation theory is presented in [1], Chapter 14.3.

The long term overall packet loss probability P_{loss} is the ratio of the average packet loss rate to the average packet arrival rate:

$$P_{loss} = \frac{1}{m}E\{(\lambda_t - c)^+\} = \frac{1}{m}\int_{\lambda_t > c}(\lambda_t - c)dP, \tag{3}$$

where m is the mean rate of the multiplexed flows, λ_t is the instantaneous arrival rate, P is the probability distribution of the instantaneous arrival rate and c denotes the link capacity. We can express P_{loss} with the instantaneous loss probability p_t as

$$P_{loss} = E\{p_t\lambda_t\}, \quad \text{where} \quad p_t = \frac{(\lambda_t - c)^+}{\lambda_t}. \tag{4}$$

Large deviation theory provides a way to approximate tail probabilities like $P\{\lambda_t > c\}$ and thus the loss probability.

First we introduce P_β, the shifted probability measure of λ_t, such that

$$dP_\beta = \frac{e^{\beta\lambda_t}}{\psi(\beta)}dP, \quad \text{where} \quad \psi(\beta) = E\{e^{\beta\lambda_t}\}, \tag{5}$$

and $\mu(\beta)$, the cumulant generating function as

$$\mu(\beta) = ln\psi(\beta). \tag{6}$$

From this, the original probability can be expressed as

$$dP = e^{-\beta\lambda_t}\psi(\beta)dP_\beta. \tag{7}$$

This shifted distribution can be accurately approximated around its mean $m(\beta) = E_\beta\{\lambda_t\}$ by a normal distribution with the same mean. Since β is a free parameter $m(\beta)$ can be moved to the value of interest for the tail probability, in our case to c. The corresponding value of β, denoted by β^* is given by

$$m(\beta^*) = c. \tag{8}$$

From the definition in Eq. 6 $m(\beta) = \mu'(\beta) = E_\beta\{\lambda_t\}$, and $\sigma^2(\beta) = m'(\beta) = \mu''(\beta)$ are the expected value and variance of the shifted distribution P_β. Consequently, if λ_t is not constant (its variance, $\sigma^2(\beta)$ is positive), then $m(\beta)$ is strictly increasing where $\mu(\lambda_t)$ is finite and equation 8 has a unique solution.

The evaluation of the integral in Eq. 3 leads to [1]

$$P_{loss} \approx \frac{1}{\sqrt{2\pi}m{\beta^*}^2\sigma(\beta^*)}e^{-\beta^*c+\mu(\beta^*)}. \tag{9}$$

The above calculated overall loss probability gives also the loss probability of the individual streams if they have the same characteristics. However, the loss distribution among streams with different characteristics will be uneven. Assume, that packets arriving in overload periods have the same loss probability independently of the source of the packets. Still, sources send a different proportion of packets during these periods. For bursty streams the bursts are correlated to overload periods as they are causing the overload themself. As a result, bursty streams experience higher loss probability than smooth ones.

The loss distribution among sources in the case of burst scale overflow can be estimated as described in [1]. Similarly to Eq. 3, the loss probability of the individual stream i is equal to

$$P_{loss}^{(i)} = \frac{1}{m_i}E\{p_t\lambda_t^{(i)}\}, \tag{10}$$

where m_i is the mean rate of the stream, $\lambda_t^{(i)}$ is the instantaneous rate, and p_t is the loss probability at time t.

The probability shift method can be applied in this case as well. Assuming, that the probability that λ_t significantly exceeds c is very small, $P_{loss}^{(i)}$ can be approximated as

$$\frac{P_{loss}^{(i)}}{P_{loss}} \approx \frac{m}{c}\frac{m_i(\beta^*)}{m_i}. \tag{11}$$

To calculate the overall loss probability using Eq. 9 and the loss distribution among the sources using Eq. 11 the value of β^*, $m_i(\beta^*)$, $\sigma_i(\beta^*)$ and $m(\beta^*)$ has to be derived. These values can be expressed in closed form if $\lambda_t^{(i)}$ and λ_t have some standard distribution (e.g., for normal distribution). In the case of real sources,however, they have to be calculated numerically. Assuming, that the distributions of the individual streams are known from measurements, and the multiplexed streams are independent, the following system of equations has to be solved

$$m(\beta^*) = c \tag{12}$$

$$m(\beta) = \sum_i m_{(i)}(\beta) \tag{13}$$

$$\sigma^2(\beta) = \sum_i \sigma_{(i)}{}^2(\beta) \tag{14}$$

$$m_{(i)}(\beta) = \frac{d}{d\beta} lnE\{e^{\beta\lambda_t^{(i)}}\} = \frac{\frac{d}{d\beta}E\{e^{\beta\lambda_t^{(i)}}\}}{E\{e^{\beta\lambda_t^{(i)}}\}} = \frac{E\{\lambda_t^{(i)}e^{\beta\lambda_t^{(i)}}\}}{E\{e^{\beta\lambda_t^{(i)}}\}} \tag{15}$$

$$\sigma_{(i)}^2(\beta) = \frac{d^2}{d\beta^2} lnE\{e^{\beta\lambda_t^{(i)}}\} = \frac{E\{\lambda_t^{(i)\,2}e^{\beta\lambda_t^{(i)}}\}E\{e^{\beta\lambda_t^{(i)}}\} - E\{\lambda_t^{(i)}e^{\beta\lambda_t^{(i)}}\}^2}{E\{e^{\beta\lambda_t^{(i)}}\}^2} \tag{16}$$

6 PERFORMANCE EVALUATION

In order to assess the effectiveness of the proposed source shaping solutions we consider the statistical quantities of the shaped video traces and packet loss statistics in case of multiplexing video streams at a single node with packet scale buffering. The presented results are based on the analytical method described in section 5 and on simulations using ns-2.

The considered scenario is shown in figure 1. It consists of n independent sources generating MPEG video streams, single buffer leaky buckets as source shapers and a multiplexing node. Each stream is shaped with some delay constraint and then multiplexed at the node with a small buffer to resolve packet scale congestion.

We present results for two MPEG-4 video traces, a soccer game with an average bit rate of 1.1 Mbps and a talk show with an average rate of 540 kbps. The traces are approximately 3600 seconds, thus 90000 frames, and 2700 seconds, thus 67000 frames long. The frames of the MPEG traces are packetized to 188 bytes, as given for the transport stream in the MPEG-2 standard [IEC61883].

Throughout the simulations we consider a single outgoing link at the multiplexing node with a capacity of 45 Mbps in the case of the soccer game and of 22.5 Mbps in the case of the talk show. We choose the link capacities proportionally to the average rate of the streams. This solution allows us to compare results at the same link utilization and level of statistical multiplexing. The multiplexing buffer can store up to 15 packets.

Trace statistics

First we consider the statistical properties of a single video stream before and after shaping for different values of maximum shaper delay d.

Figures 2 and 3 show the number of transmitted bits in a frame time for the original and shaped trace of the soccer game and the talk show respectively. The maximum shaper delays are 40 ms and 120 ms, the shaping is performed using the BC shaper algorithm. Even the relatively small shaper delay of 40 ms allows a significant reduction of the rate fluctuation. The maximum transmission rate is decreased from 3.6Mbps to 3Mbps for the soccer game trace and from 3.1Mbps to 2Mbps for the talk show trace.

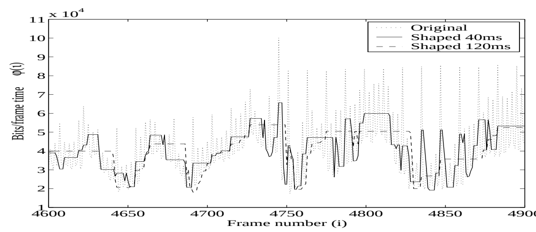

Figure 2: Number of bits transmitted in a frame time for the soccer game trace without shaping and with shaping for d=40 ms and d=120 ms, using the BC shaper algorithm.

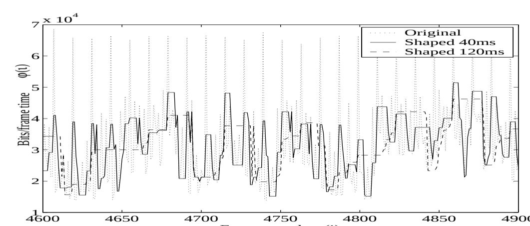

Figure 3: Number of bits transmitted in a frame time for the talk show trace without shaping and with shaping for d=40 ms and d=120 ms, using the BC shaper algorithm.

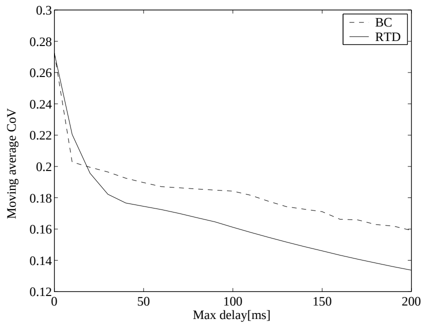

Figure 4: CoV of the shaped soccer trace versus maximum shaper delay d, considering the BC and RTD shaping algorithms.

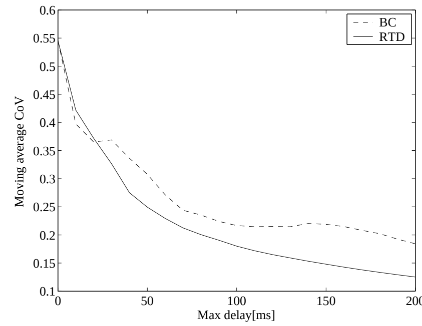

Figure 5: CoV of the shaped talk show trace versus maximum shaper delay d, considering the BC and RTD shaping algorithms.

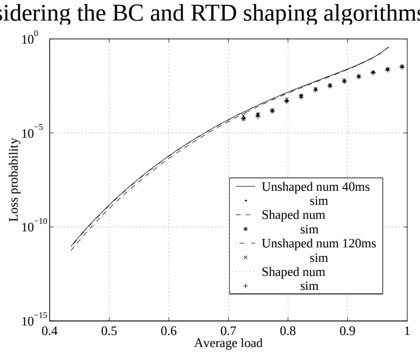

Figure 6: Loss probability of shaped and unshaped streams for 1 shaped stream. The talk show trace and the BC shaping algorithm is considered. Simulation validates the analytical results.

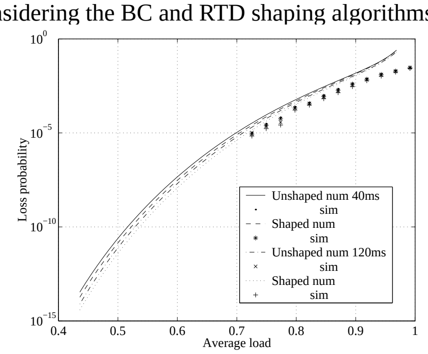

Figure 7: Loss probability of shaped and unshaped streams for half of the streams shaped. The talk show trace and the BC shaping algorithm is considered. Simulation validates the analytical results.

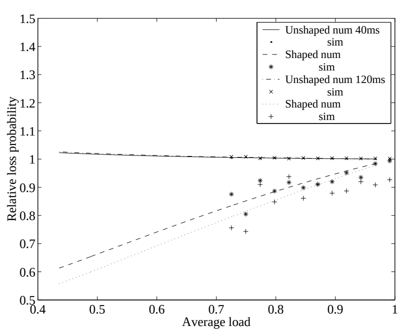

Figure 8: Relative loss probability of shaped and unshaped streams for 1 shaped stream. The talk show trace and the BC shaping algorithm is considered. Simulation validates the analytical results.

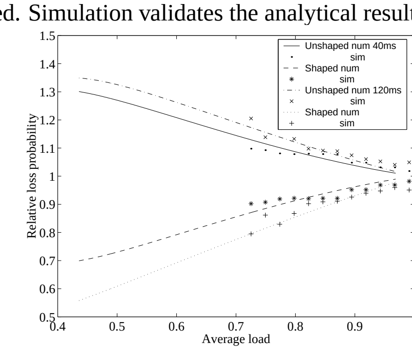

Figure 9: Relative loss probability of shaped and unshaped streams for and half of the streams shaped. The talk show trace and the BC shaping algorithm is considered. Simulation validates the analytical results.

Figures 4 and 5 show the coefficient of variation (CoV) of the traces, defined as

$$CoV = \sqrt{\frac{1}{N_{tr}} * \sum_{i=1}^{N_{tr}} (\varphi(i) - \frac{\sum_{j=i-N_{GOP}}^{i-1} \varphi(j)}{N_{GOP}})^2 * \frac{1}{E\{\varphi\}}} \quad (17)$$

where N_{tr} is the number of frames in the trace, $\varphi(i)$ is the number of bits transmitted in the ith frame time and $E\{\varphi\} = E\{f\}$ is the average number of bits transmitted in one frame time. The CoV is calculated using the moving average over one GOP time as mean value. This way the CoV reflects the frame to frame rate fluctuations without the rate variation due to the scene changes in the trace. The two curves in the figure show the CoV values for the two proposed shaping solutions.

The graphs show that for small values of shaper delay d the CoV decreases very rapidly, reflecting that a delay of a couple of frame times (20-80 ms) allows the smoothing of the transmission rates of consecutive I, P and B frames. At larger delays the marginal gain decreases significantly.

The two shaping methods result in similar changes in the CoV values. As expected, the RTD method decreases the rate fluctuations better. For large values of d the difference is around 20% in the terms of CoV reduction, since the RTD method can adjust the shaper rate more precisely based on the information maintained in the buffer, while the BC method over- and underestimates the shaper rate and has to make corrections later. As the difference is not significant, due to its simplicity we focus on the BC method in the followings.

Packet loss probabilities

In this part we investigate the average packet loss probability of multiplexed video streams as a function of the average load, defined by the ratio of the sum of the mean rates of the streams to the link transmission capacity. The presented results are based on mathematical analysis, simulation results are shown to demonstrate the accuracy of the method. Simulations were run 20000 to 100000 seconds to have enough loss events even in the case of loss probabilities in the order 10^{-5}.

We show results for shaping the talk show trace with the BC method. To see the effect of introducing shaping gradually at the sources two scenarios are considered. In the first one only one stream is shaped, while all the other multiplexed streams are transmitted unshaped. In the second scenario half of the streams are shaped at the source.

Figure 6 shows the loss probability of the shaped and unshaped sources for shaper delays of 40 ms and 120 ms in the case of 1 shaped source. Figure 7 shows the results for the scenario where half of the sources are shaped. Figures 8 and 9 show the relative loss probabilities of the shaped and unshaped sources compared to the average loss probability.

The numerical results are validated by simulation. The results reflect, that the mathematical analysis works well for losses up to 10^{-2}, then slightly overestimates the loss probability as a consequence of the large deviation approximation.

Comparing figures 6 and 7 it can be seen that shaping half of the sources decreases the overall loss probability by roughly one order of magnitude. The results with different shaper delays show that in the case of the considered talk show trace, shaping with a delay of 40 ms is almost as efficient as shaping with a delay of 120 ms in the terms of reducing the loss probability. For the soccer trace shaping with a delay of 120 ms has a slightly bigger effect. This is due to the lower ratio of temporal redundacy which induces lower peak to mean ratio.

The results show in Figures 8 and 9 that the decrease in the loss probablity achieved by shaping the sources increases as the average load, and thus the loss rate decreases. Since the desired loss probability of video streams is in the order 10^{-5}, the difference can be up to 35%, even if only one stream is shaped. The gain achieved depends on the stream characteristics, for the soccer trace the experienced gain was less, around 20% at a loss probability around 10^{-5}.

Packet loss probabilities in I, P and B frames

In addition to the average packet loss probability of the streams it is worthwhile to evaluate the packet loss probability in individual frame types, since it affects the percepted visual quality.

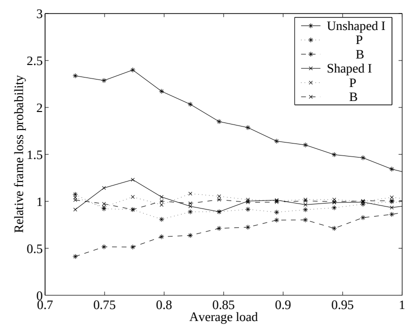

Figure 10: Relative packet loss probability in I,P,B frames of the shaped and unshaped streams for d = 120 ms, half of the streams shaped. The talk show trace and the RTD shaping algorithm is considered. Simulation results.

Figure 10 shows the packet loss probability in I, P and B frames relative to the average loss probability for the shaped and unshaped sources for the scenario when half of the sources is shaped with a maximum shaper delay of 120 ms. The figure shows that while in the case of unshaped sources the loss probability in the I frames is the highest, up to 100% above the average loss probability and that in the B frames is the lowest, in the case of shaped sources the loss probabilities in the individual frame types are roughly the same. In the I frames the decrease of loss probability is around 60%. Consequently, as losses in the I frame have a significant effect on the visual quality, the positive effects of the shaping include not only lower loss probability but also the improved distribution of these losses among the frame types.

The presented results show the following effects of source shaping. Considering the trace statistics, shaper delays in the 20-40 ms range decrease the CoV of the stream significantly, higher delays introduce decreasing marginal gains. Comparing the two proposed shaping algorithms, the simple BC algorithm works rather well, especially at small shaper delays. Results on multiplexing the video streams at a multiplexer with small buffer show that the shaped streams experience lower loss probabilities than the unshaped ones, the difference is about 30%. The gradual introduction of the source shaping in the network has a significant effect, the loss probabilities decrease with one order of magnitude if half of the sources adopt shaping. The positive effect of shaping is reflected by the distribution of losses among the different frame types in the video stream. Multiplexing unshaped streams results high loss probability for the I frames, this loss probability decreases significantly if the stream is shaped.

7 CONCLUSION

In this paper we proposed and evaluated solutions that assist live video transmission over the Internet. Specifically, we considered the scenario, when the MPEG coded video streams are shaped at the source host and network routers provide small buffers to resolve packet scale congestion, motivated by the current trends of designing traffic control solutions, where the main idea is to add functions to the hosts and keep the operation of the network routers simple.

We proposed computationally simple shaping algorithms that are adequate to shape live video streams with a wide range of delay tolerance at the shaper and provided analytical method to evaluate the efficiency of the proposed solutions.

The analytical and simulation-based performance evaluation proved that i) a simple shaper algorithm based on buffer occupancy and delay limit results in efficient shaping in many cases; ii) even shaping with very low delay bound, adequate for real-time applications, improves the performance in terms of packet loss probability and the packet loss distribution among the different frame types; and iii) shaping provides a means to improve the quality of individual video transmissions even if not all the hosts shape their traffic.

Finally, as ongoing work we further investigate the methods to assist video transmission over the Internet. Specifically, we are interested in the efficiency of source shaping versus buffering at the routers and buffering versus forward error correction.

References

[1] "Broadband network teletraffic, Final report of action COST 242," Springer, 1996.

[2] G. Bianchi, A. Capone, Ch. Pertioli, "Throughput Analysis of End–to–End Measurement-Based Admission Control in IP," in *Proc. of IEEE INFOCOM 2000,*, March 26–30, 2000, pp. 1461–1470.

[3] L. Breslau, E. W. Knightly, S. Shenker, I. Stoica and H. Zhang, "Endpoint Admission Control: Architectural Issues and Performance," in *Proc. of ACM SIGCOMM 2000*, 28 Aug. – 1. Sept., 2000, pp. 57-69.

[4] R. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Transactions on Information Theory,* vol. 37, no. 1, Jan. 1991, pp. 114-131.

[5] R. B. Gibbens and F. P. Kelly, "Distributed Connection Acceptance Control for a Connectionless Network," in *Proc. of the 16th International Teletraffic Congress*, June 7–11, 1999, pp. 941-952.

[6] V. Elek, G. Karlsson, R. Ronngren, "Admission Control Based on End-to-end Measurements," in *Proc. of IEEE INFOCOM 2000,*, March 26–30, 2000, pp. 623–630.

[7] E. W. Knightly, "H-BIND: a New Approach to Providing Statistical Performance Guarantees to VBR Traffic," *IEEE Infocom'96,* March 1996.

[8] E. W. Knightly and H. Zhang, "D-BIND: an Accurate Traffic Model for Providing QoS Guarantees to VBR Traffic," *IEEE/ACM Transactions on Networking,* vol.5, no.2, April 1997, pp.219-231.

[9] H. Zhang and E. W. Knightly, "RED-VBR: A Renegotiation-based Approach to Support Delay-Sensitive VBR Video," *ACM Multimedia system Journal,* May 1997.

[10] Simon S. Lam, Simon Chow, David K. Y. Yau,"An Algorithm for Lossless Smoothing of MPEG video," *ACM SIGCOMM Computer Communication Review,* vol. 24, no. 4, Oct. 1994, pp. 281-293

[11] J-Y. Le Boudec and O. Verscheure, "Optimal Smoothing for Guaranteed Service," *IEEE Transactions on Networking,* vol.8, no.10, Dec. 2000.

[12] N. Modani, P. Dube and A. Kumar, "Measurement Based Optimal Source Shaping with a Shaping+Multiplexing Delay Constraint" *Proc. of IEEE Infocom 2000*, March 26–30, 2000,pp. 1807-1816.

[13] J. Rexford, S. Sen, J Dey, W. Feng, J. Kurose, J. Stankovic and D. Towsley, "Online Smoothing of Live, Variable-bit-rate Video" *Proc. of International Workshop on Network and Operating Systems Support for Digital Audio and Video*, May 1997, pp. 249-257

[14] J. W. Roberts, "Traffic Theory and the Internet," *IEEE Communications Magazine,* Jan. 2001., pp.94-99.

[15] M. Reisslein, K. W. Ross, S. Rajagopal, "Guaranteeing Statistical QoS to Regulated Traffic: The Multiple Node Case," in Proc. *IEEE Decision & Control'98,* pp. 531-538, 1998.

[16] M. Reisslein, K. W. Ross, S. Rajagopal, "Guaranteeing Statistical QoS to Regulated Traffic: The Single Node Case," in Proc. *IEEE Infocom'99,* pp. 1061-1072, 1999.

[17] A. Mankin, F. Baker, B. Braden, S. Bradner, M. O'dell, A. Romanow, A. Weiurib, and L. Zhang, "Resource ReSerVation Protocol - Version 1 Applicability Statement Some Guidelines on Deployment. RFC 2208," September 1997.